

Sounds@Work - Akustische Repräsentationen für die Mensch-Computer Interaktion in kooperativen und hybriden Arbeitsumgebungen

Vom Fachbereich 18
Elektrotechnik und Informationstechnik
der Technischen Universität Darmstadt
zur Erlangung der Würde
eines Doktor-Ingenieurs (Dr.-Ing.)
genehmigte Dissertation

von

Christian Müller-Tomfelde
(Diplom-Ingenieur)
geboren am 24.2.1965
in Hamburg

Referent:	Prof. Dr.-Ing. Ralf Steinmetz
Korreferent:	Prof. Dr.-Ing. Dr. techn. h.c. Jens Blauert
Tag der Einreichung:	18.11.2002
Tag der mündlichen Prüfung:	21.1.2003

D17
Darmstädter Dissertation

Danksagung

Herrn Prof. Dr.-Ing. Ralf Steinmetz gilt mein Dank für die Initiierung und unterstützende Begleitung dieser Arbeit. Herrn Prof. Dr.-Ing. Dr. techn. h.c. Jens Blauert danke ich für die Übernahme des Korreferats sowie Herrn Prof. Dr.-Ing. Bernhard Dorsch für den Vorsitz des Prüfungsausschusses. Zu großem Dank verpflichtet bin ich Dr. rer. nat. Dr. phil. Norbert A. Streitz für seine Unterstützung und Zusammenarbeit im i-LAND Projekt des von ihm geleiteten Forschungsbereichs „AMBIENTE - Arbeitswelten der Zukunft“ am Fraunhofer Institut für Integrierte Publikations- und Informationssysteme (IPSI) in Darmstadt. Weitere Unterstützung wurde mir zuteil durch die Mitarbeiter des Bereichs AMBIENTE Jörg Geißler, Peter Tandler, Thorsten Holmer und Shin'ichi Konomi. Bei Sascha Steiner und Tobias Münch bedanke ich mich für ihren überdurchschnittlichen Einsatz und Fleiß bei der Mitarbeit im Projekt. Wolfgang Reischl und Daniel Warth möchte ich für ihr besonderes Engagement bei der Gestaltung der Roomware-Komponenten der ersten Generation danken. Ihnen und allen anderen, wissenschaftlichen und studentischen Mitarbeitern des Bereichs AMBIENTE möchte ich für ihre Unterstützung danken. Herzlichster Dank an Uta Daur für die sorgfältige Durchsicht der Arbeit. Ganz herzlich danke ich meinen Eltern für die liebevoll unterstützende Begleitung meines bisherigen Lebenswegs.

Inhaltsverzeichnis

1	Einleitung	1
2	Mensch-Computer Interaktion	5
2.1	Kommunikationsmodelle	6
2.2	Interaktionsmodelle	8
2.2.1	Wahrnehmung	11
2.2.2	Multimodale Interaktion.....	14
2.3	Die Rolle des Computers.....	16
2.3.1	Interaktion in virtuellen Umgebungen	18
2.3.2	Ubiquitous Computing	20
2.4	Akustische Signale bei der Mensch-Computer Interaktion.....	22
3	Grundlagen der akustischen Wahrnehmung.....	25
3.1	Psychoakustik.....	26
3.1.1	Komponenten der auditiven Wahrnehmung	26
3.1.2	Räumliches Hören.....	28
3.1.3	Effekte auditiver Wahrnehmung	29
3.2	Ecological Acoustics	32
3.2.1	Invariante Objektstrukturen	34
3.2.2	Elementare akustische Ereignisse	35
4	Digitale Signalverarbeitung in Computeranwendungen	39
4.1	Audiosignalverarbeitung in Echtzeit	41
4.2	Verzögerungen der Datenströme	43
4.2.1	Blockverarbeitung	43
4.2.2	Behandlung von Ereignissen.....	45
4.3	Erzeugung von Audiosignalen	45
4.3.1	Syntheseverfahren	46
4.3.2	Designverfahren	48

5	Neue Ansätze akustischer Repräsentationen in hybriden Umgebungen	51
5.1	Hybride Umgebungen	51
5.2	Forderung nach hörbaren Objekteigenschaften.....	53
5.3	Akustisches Verhalten von hybriden Objekten	55
5.4	Technische Aspekte hybrider Umgebungen.....	56
6	Realisierungen von akustischen Repräsentationen für die i-LAND Umgebung	59
6.1	Einordnung akustischer Repräsentationen	60
6.2	Gestaltungsregeln für akustische Darstellungen	62
6.3	Unterstützung von Interaktion an der DynaWall	63
6.3.1	Akustische Darstellung für die Interaktion an der DynaWall	65
6.3.2	Virtuelle Reibung	66
6.3.3	Akustische Unterstützung von Gesteninteraktion	67
6.3.4	Interaktion und Kooperation mit der DynaWall	70
6.3.5	Erweiterung der technischen Umgebung der DynaWall.....	71
6.4	Erweiterte akustische Modellierung von Oberflächeninteraktionen	77
6.4.1	Modell der Stiftinteraktion	78
6.4.2	Modell passiver akustischer Objekte	79
6.4.3	Stimulation des akustischen Systems.....	81
6.4.4	Akustische Texturen von Oberflächen.....	81
6.4.5	Geräuschgenerator.....	82
6.4.6	Testumgebung für akustische Oberflächentexturen.....	84
6.5	Passage – Physikalisches ‚Lesezeichen‘ in die virtuelle Welt	85
6.5.1	Anwendung und Einordnung des Passage-Mechanismus.....	86
6.5.2	Architektur für den Passage-Mechanismus.....	87
6.5.3	Unterstützung der Interaktion durch akustisches Feedback.....	90
6.6	Akustische Darstellung von Zuständen in Computersystemen.....	92
6.6.1	Der geräuschlose Computer	92
6.6.2	Der akustische Systemmonitor.....	94
6.6.3	Klanggenerator für Systemaktivitäten.....	95
6.7	Hybride Präsentation akustischer Signale	96
6.7.1	Kontinuum der Präsentation akustischer Darstellungen	97

6.7.2	Architektur der hybriden akustischen Darstellung.....	98
6.7.3	Anwendungen für hybride akustische Darstellungen	101
7	Entwicklungen zur Echtzeit-Signalverarbeitung auf Standardcomputern	107
7.1	Bibliothek von Objekten zur Signalverarbeitung.....	107
7.1.1	Integration in Anwendungen.....	109
7.1.2	Objektgruppen.....	110
7.1.3	Makro-Umgebung zur standardisierten Programmierung	114
7.2	Beispiele zur Signalgeneration und -manipulation	118
7.2.1	Physikalische Modelle	118
7.2.2	Abstrakte Modelle.....	119
7.3	Leistungsverhalten von Standardcomputersystemen	120
7.3.1	Schnelle Faltung als Testoperation	121
7.3.2	Speichermodelle.....	122
7.3.3	Hardware-Plattformen und Betriebssysteme.....	126
7.3.4	Prozessorspezifische Optimierung	128
7.3.5	Ergebnisse der Leistungsvergleiche.....	128
7.4	Signalverarbeitung in parallelen Prozessen.....	130
7.4.1	Beispiel einer Implementierung eines Faltungsalgorithmus	131
7.4.2	Mechanismus der Verarbeitung	132
7.4.3	Gesamtsystem der Verarbeitung	134
7.4.4	Ergebnisse der Implementierung.....	135
8	Zusammenfassung	139
	Literaturverzeichnis	141
	Anhang	151
A.1	Inhalt der Daten-CD	151
A.2	Abkürzungen	155
A.3	Publikationsliste im Kontext der Dissertation.....	157
A.4	Lebenslauf	159

1 Einleitung

Die Interaktion des Menschen mit Objekten seiner Umgebung ist immer auch begleitet durch Klänge und Geräusche, die bei dieser Interaktion entstehen. Die Klänge und Geräusche helfen dem Menschen Zustandsveränderungen der Objekte wahrzunehmen und weitere Möglichkeiten zur Interaktion mit dem Objekt zu entdecken. Technische Entwicklungen und Verbesserungen haben in vielen Fällen auch zur Folge, dass die bei der Interaktion mit technischen Geräten vertrauten akustischen Eigenschaften der Objekte verloren gehen. Da dieser Verlust aber zu Irritationen und Missverständnissen bei der Interaktion führen kann, werden entsprechende Geräusche oft künstlich wieder hinzugefügt, wie z. B. im Fall des Kfz-Blinkerrelais. Auch im Bereich der Informationstechnologie kann z. B. durch die Miniaturisierung der Komponenten und dem damit verbundenen ‚Verschwinden‘ der Geräte (*disappearing computer*) in die Umgebung der Verlust von computertypischen Geräuschen beobachtet werden. Das Wiederhinzufügen von Klängen und Geräuschen bei der Interaktion mit Computern kann nun einerseits dazu genutzt werden, vorhandene Erwartungen der Benutzer zu erfüllen, andererseits aber auch neuartige virtuelle akustische Eigenschaften zu erzeugen. Diese Ausgangssituation eröffnet neue Möglichkeiten und Herausforderungen für den Einsatz von akustischen Repräsentationen, insbesondere für die Interaktion und Kooperation in computerunterstützten Arbeitsumgebungen.

Ziel dieser Arbeit ist es, an Hand einer computerunterstützten Umgebung für kooperatives Arbeiten exemplarisch aufzuzeigen, wie akustische Repräsentationen realisiert und genutzt werden können, um die Interaktion von Menschen in zukünftigen *hybriden Umgebungen* zu unterstützen. Jenseits von traditionellen Desktop-Computern und virtuellen Umgebungen sollen Umgebungen als *hybrid* bezeichnet werden, wenn sie reale und virtuelle Anteile integrieren und dadurch neue Möglichkeiten der Interaktion schaffen. Der Einsatz von akustischen Repräsentationen erfolgt dabei in einer neuen Form vornehmlich als *non-speech audio*. Dies ist in mehrfacher Hinsicht abzugrenzen von anderen Verwendungsarten akustischer Signale. So erfolgt die Verwendung des akustischen Mediums weder als zusätzlicher Kommunikationskanal, wie z. B. der Tonkanal zum Bild in einem Video, noch soll dem Benutzer mithilfe einer akustischen Sinnestäuschung das Eintauchen in eine computererzeugte Illusion als eine ‚alternative Realität‘ erleichtert werden. In dieser Arbeit soll auch die Interaktion zwischen Mensch und Computer mithilfe von Sprache keine Berücksichtigung finden. Sprache soll der sozialen Interaktion zwischen den Benutzern der Arbeitsumgebung vorbehalten bleiben.

Die Arbeit stellt vielmehr neue Konzepte und Realisierungen der Anreicherung von Interaktion mit akustischen Repräsentationen (*audio augmentation*) vor, die die hybriden Qualitäten der Umgebung ausnutzen, unterstreichen und erweitern. Die Beispiele und Prototypen in dieser Arbeit belegen, wie auf technologische Entwicklungen angemessen reagiert werden kann, um kohärente und plausible Interaktionen in zukünftigen hybriden Arbeitsumgebungen zu erzeugen.

Seit dem Aufschwung im Bereich der Multimedia-Technologie Ende der 80er Jahre gehört die Möglichkeit der Ausgabe von Audiosignalen zur Standardausstattung von Computern. Von dieser Zeit an wurde auch die Forschung im Bereich der Verwendung akustischer Signale für die Interaktion stark intensiviert. So erschien 1989 eine Sonderausgabe der Fachzeitschrift *Human-Computer Interaction* über *non-speech audio*, in der z. B. eine Erweiterung der Computer-Desktopmetapher mit *Auditory Icons* vorgestellt wurde (*SonicFinder*, Gaver 1989). Seitdem wurden in verschiedenen Arbeiten die Verwendung und der Nutzen von akustischen Repräsentationen für die Interaktion zwischen Mensch und Computer untersucht. Mitte der 90er Jahre gab die Entwicklung im Bereich dreidimensionaler akustischer Projektion der Forschung über akustische Darstellungen zusätzliche Impulse.

Bis auf wenige Ausnahmen behandeln diese Arbeiten aber nur die Unterstützung von Einzelpersonen und berücksichtigen auch nicht die Entwicklungen integrierter und mobiler Technologien in Arbeitsumgebungen für Gruppen. Für zukünftige hybride Arbeitsumgebungen, die nicht nur aus mehreren vernetzten Standardcomputern bestehen oder nicht ausschließlich in einer virtuellen Umgebung realisiert werden, sind neuartige akustische Darstellungen erforderlich. Sie müssen gleichzeitig im realen wie im virtuellen Raum eine kohärente Wahrnehmung ermöglichen, um damit die Interaktion mit dem Computer und die Kooperation in der Gruppe zu unterstützen.

Um dies zu erreichen, können z. B. Ansätze der Klangmodellierung für die Interaktion mit einem Stift auf einer Oberfläche genutzt werden, um verschiedene Charakteristika zu vermitteln. Dabei können dem Benutzer Zustände von Objekten und andere Informationen bei der Interaktion an der Oberfläche akustisch dargestellt werden. Das Zusammenfallen der taktilen Aktion an realen Oberflächen mit einer akustischen Darstellung virtueller Eigenschaften eines Artefakts führt zu einer besonderen Situation: Das Artefakt, mit dem der Benutzer interagiert, kann als hybrid bezeichnet werden, da es dem Benutzer bei der Interaktion sowohl reale als auch virtuelle Qualitäten zur Wahrnehmung anbietet.

Die vorliegende Arbeit besitzt zwei inhaltliche Schwerpunkte: die Anreicherung der Mensch-Computer Interaktion mit akustischen Repräsentationen und die digitale

Audiosignalverarbeitung mithilfe von Standardcomputern. Mit dieser Arbeit sind neue Ansätze der akustischen Repräsentationen für die Mensch-Computer Interaktion (*Sounds@Work*) entwickelt worden. Auf der Grundlage dieser Ansätze entstanden im Rahmen dieser Arbeit Prototypen für die Roomware-Komponenten der i-LAND Umgebung am Fraunhofer Institut für Integrierte Publikations- und Informationssysteme (IPSI) in Darmstadt.

Die Arbeit ist wie folgt gegliedert: Nach dieser Einleitung (Kapitel 1) werden im zweiten Kapitel die Grundlagen der Mensch-Computer Interaktion vorgestellt und die Bedeutung des Begriffs Kommunikation über seine technische Erklärung hinaus erörtert. Dabei werden neue Konzepte der Interaktion für *Ubiquitous Computing* und die Interaktion in virtuellen Umgebungen erläutert. Eine Zusammenstellung der wichtigsten Arbeiten im Bereich akustischer Darstellungen schließt das Kapitel ab.

Im dritten Kapitel werden allgemeine Aspekte und grundlegende Komponenten der auditiven Wahrnehmung vorgestellt. Für die Verwendung akustischer Signale bei der Mensch-Computer Interaktion soll hier der Schwerpunkt auf sog. *non-speech audio* liegen. Diskutiert werden in diesem Kapitel außerdem spezielle Ansätze aus dem Bereich der Wahrnehmungspsychologie und deren Bedeutung für die Mensch-Computer Interaktion.

Das vierte Kapitel beschäftigt sich mit technischen Grundlagen der digitalen Audiosignalverarbeitung insbesondere vor dem Hintergrund der Verwendung auf Standardcomputern und es werden wichtige Eigenschaften und Verfahrensweisen zur Erzeugung von Audiosignalen vorgestellt.

Im fünften Kapitel wird herausgestellt, warum die akustischen Darstellungen für kooperative Situationen in hybriden Umgebungen besondere Aufmerksamkeit erfordern. Jenseits der speziellen und teilweise eingeschränkten Möglichkeiten von traditionellen Computerumgebungen und virtuellen Umgebungen muss für die akustischen Repräsentationen in hybriden Umgebungen immer die Balance zwischen dem realen Objekt und seinen erweiterten Möglichkeiten gehalten werden. Außerdem wird die Tatsache berücksichtigt, dass Entwicklungen in der Computertechnologie zu immer kleineren und leiseren bzw. geräuschlosen Geräten führen. Ähnlich wie ein optisches Display eine minimale Größe nicht unterschreiten darf, um *Lesbarkeit* zu ermöglichen, wird daher für die akustischen Darstellungen u. a. die Forderung nach *hörbaren Objekteigenschaften* aufgestellt. Beispielhafte Realisierungen dieser Darstellungen werden dann ausführlich in Kapitel 6 beschrieben. Sie erfolgten für die Roomware-Komponenten der hybriden i-LAND Umgebung und wurden im Bereich AMBIENTE (Arbeitswelten der Zukunft) am Fraunhofer-IPSI durchgeführt.

Im abschließenden Kapitel 7 werden technische Details vorgestellt, die sich mit den Möglichkeiten der Erzeugung und Manipulation von Audiosignalen in Echtzeit auf Standardcomputern beschäftigen. Hierfür wurde eine Bibliothek von Signalobjekten für die Verarbeitung in Echtzeit geschaffen, die zur Erzeugung der verschiedenen Anwendungen verwendet wurde. Beispiele dafür werden detailliert beschrieben. Für einen besonderen Fall der Faltung von Audiosignalen mit Impulsantworten mit einer großen Anzahl von Filterkoeffizienten wurde ein Algorithmus auf einem Standardcomputer implementiert, der in Echtzeit eine geringe Ausgabeverzögerung aufweist. Es wird damit demonstriert, auf welche Weise Standardcomputer mit speziellen Verfahren für die Erzeugung von akustischen Darstellungen bei der Mensch-Computer Interaktion in Echtzeit geeignet sind.

Um die Frage nach der Entwicklung und Realisierung von akustischen Repräsentationen für die Unterstützung der Mensch-Computer Interaktion adäquat zu beantworten, ist es notwendig, Erkenntnisse aus verschiedenen Disziplinen (Elektrotechnik, Informatik, Wahrnehmungspsychologie, Psychoakustik) zu berücksichtigen. Die Herausforderung, dieser Arbeit besteht darin, neue Konzepte der Mensch-Computer Interaktion und neue technologische Möglichkeiten im Hinblick auf das Ziel der Realisierung neuartiger Ansätze akustischer Repräsentationen für die Interaktion in hybriden Umgebungen zusammenzuführen.

2 Mensch-Computer Interaktion

Akustische Repräsentationen gehören nach optischen Darstellungen zu den wichtigsten, aber auch häufig ignorierten Bestandteilen bei der Interaktion zwischen Mensch und Computer. Welche Bedeutung den akustischen Signalen bei der Interaktion zukommt, wird oft ersichtlich, wenn ein akustisches Signal nicht mehr ertönt. Die vertraute Handlung, die trotz der Stille zum gleichen Ergebnis führt, kann auf den Benutzer irritierend wirken, wenn der akustische Bestandteil der Interaktion fehlt. Um eine Grundlage für die Entwicklung von akustischen Repräsentationen für die Mensch-Computer Interaktion zu schaffen, soll in diesem Kapitel eine Darstellung und Diskussion der Bereiche der Mensch-Computer Interaktion erfolgen, die hierfür relevant sind. Das Verständnis für die Bedeutung der Information, der Kommunikation, der menschlichen Wahrnehmung aber auch der Rolle des Computers bei der Interaktion über technische Kenntnisse hinaus ist entscheidend, um akustische Repräsentationen entwickeln zu können.

Akustische Signale dienen einerseits der Wahrnehmung von Ereignissen und der Orientierung in der Umgebung, andererseits dem Austausch von Informationen zwischen Menschen. Interaktion soll hier als Wechselbeziehung verstanden werden, als Handlung oder Tätigkeit zwischen mehreren Menschen oder Menschen und Gegenständen bzw. der Umgebung. Unter Kommunikation ist ein spezieller Teil der Interaktion zu verstehen, in dem Informationen bzw. Nachrichten in expliziter Form von Sprache oder Schrift ausgetauscht werden. In manchen Anwendungsbereichen der Kommunikation, wie z. B. der Telefonie, dominieren dann die Eigenschaften der hörbaren Signale über denen der sichtbaren. Im besonderen Fall der computervermittelten Kommunikation erlauben multimediale Technologien eine Bereicherung der Mensch-Mensch Interaktion und Kommunikation über die verschiedenen räumlichen, zeitlichen und sprachlichen Grenzen hinweg. Der Begriff Mensch-Computer Kommunikation soll hier vermieden werden, da seine Bedeutung umfassender ist, als es im Verhältnis Mensch und Computer möglich ist (Herrmann 1986). In der vorliegenden Arbeit soll die Verwendung von akustischen Signalen bei der Mensch-Computer Interaktion im Vordergrund stehen. Insbesondere sollen die Situationen, in denen der Mensch als Empfänger von akustischen Signalen betrachtet wird, detailliert erörtert werden, da dies für einen Großteil der Anwendungssituationen im Kontext dieser Arbeit der Fall sein wird. Ausgehend von der grundlegenden Kombination Mensch-Computer soll auch die computervermittelte und -unterstützte Interaktion und Kommunikation betrachtet werden, bei der eine Gruppe mehrerer Menschen mithilfe von Computern untereinander kommunizieren und kooperieren

kann. Mit dem Begriff der Umgebung oder Umwelt, wie in Abb. 1 dargestellt, sollen alle Teile zusammengefasst werden, die nicht genauer für die Interaktion spezifiziert werden können wie z. B. Arbeitswelt und Organisation. Daneben kann ‚Umgebung‘ den Hintergrund bezeichnen, vor dem alle anderen Interaktionen stattfinden (Kontext).

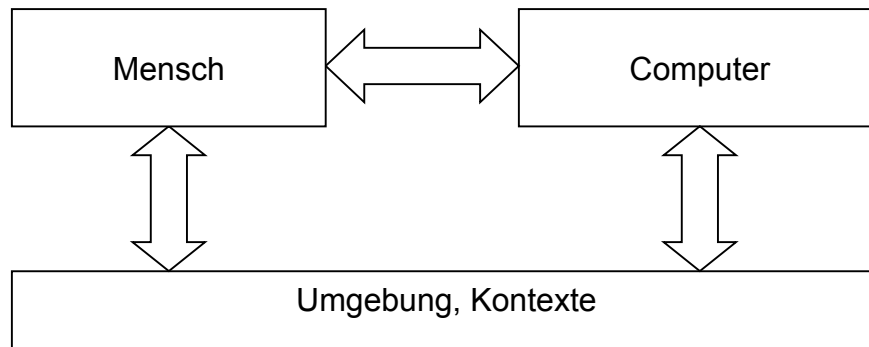


Abb. 1: Möglichkeiten der Interaktion zwischen Mensch, Computer und Umgebung bzw. Kontext.

Weiterhin ergeben sich bei spezieller Betrachtung besondere Formen der Interaktion: Wird z. B. der Computer in die Umgebung integriert, wird für den Menschen dadurch die Möglichkeit zur Interaktion mit dem Computer „allgegenwärtig“ (Weiser 1991). Wird hingegen der Mensch von der natürlichen Umgebung getrennt und interagiert er ausschließlich mit dem Computer, kann für den Benutzer eine sog. virtuelle Realität entstehen. Die neutrale Anordnung, wie sie in Abb. 1 dargestellt, ist beispielsweise in der Situation mit einem herkömmliche Desktop-Computer gegeben.

2.1 Kommunikationsmodelle

Traditionell wird in den Fachgebieten der Ingenieurwissenschaften, insbesondere der Nachrichtentechnik, das Kommunikationsmodell nach Shannon und Weaver von 1949 herangezogen, wenn es z. B. darum geht, die Eigenschaften einer Übertragungsstrecke bzw. eines Übertragungskanal zu messen oder den Einfluss von Signalkodierungen zu bewerten. In der mathematischen Theorie der Kommunikation schlägt Shannon vor, den Informationsgehalt eines Zeichens über die logarithmische Abbildung der Wahrscheinlichkeit des Auftretens des Zeichens in einer bestimmten Situation zu definieren. Im extremen Fall bedeutet dies, dass ein Informationsmaximum für ein Zeichen erreicht wird, wenn alle Zeichen gleich wahrscheinlich sind. Bezogen auf ein analoges Kanalsignal muss dann ein sog. weißes Rauschen aufgrund seiner spektralen Gleichverteilung maximale Information besitzen. Bei gegebener Kanalbandbreite und unter der Bedingung einer maximalen Kanalauslastung ist dies z. B. im Bereich der

statistischen Nachrichtentechnik die Grundlage für Optimierungsmethoden wie die der linearen Prädiktion.

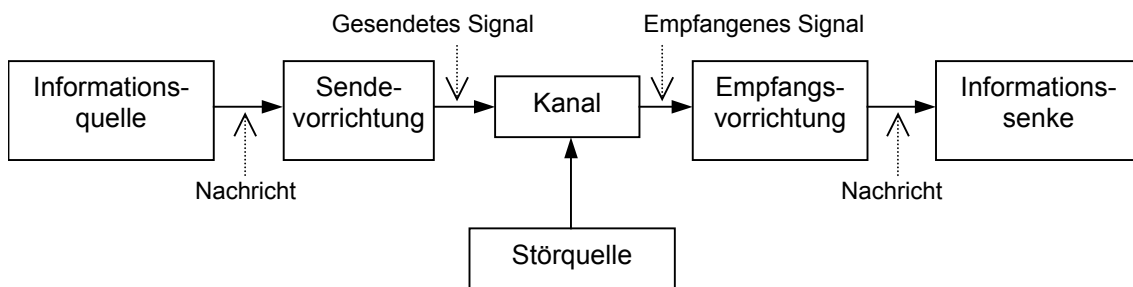


Abb. 2: Kommunikationsmodell nach Shannon und Weaver (1949).

Vor dem Hintergrund der Kommunikation zwischen Menschen erscheint es jedoch widersinnig, dass Rauschen maximale Information besitzen soll. Hier wird deutlich, dass die Gültigkeit des Modells nach Shannon und Weaver für eine Verallgemeinerung der Begriffe Information und Kommunikation Probleme aufwirft. Das Modell geht von einer starren Konstellation von Sender und Empfänger aus, die beide über die Dauer der Kommunikation ein unverändertes Verhalten aufweisen. Zum anderen kann sich der Übertragungskanal verändern oder wie in Abb. 2 dargestellt, ein Störsignal dem eigentlichen Nachrichtensignal überlagern. Solche Eigenschaften sind überaus typisch für technische Kommunikationsanwendungen, bei denen die Endgeräte nach festgelegten Regeln Zeichen kodieren, während der Übertragungskanal Störungen und Verzerrungen unterworfen sein kann. Zu den historischen Beispielen technischer Nachrichtenübermittlung mit derartigen Eigenschaften gehört der optische Telegraph (1794, Claude Chappe) und sein Nachfolger, der elektrische Telegraph (1837, Samuel Morse).

Ernst von Weizsäcker (1974) erweitert den Begriff der Information, indem er seiner Beschreibung der Information die Eigenschaften der Erstmaligkeit und Bestätigung eines Ereignisses durch den Empfänger hinzufügt. Diese Erweiterung nennt er pragmatische Information. Er geht dabei von der Veränderbarkeit des Empfängers aus, der sich durch die empfangene Information wandeln kann (pragmatisch). Der Empfänger wird hier als lernfähiges System angesehen, das sich über die Zeit hinweg verändert. Statt der Wahrscheinlichkeit der technischen Information werden die Vielfalt und die Inkongruität der Information betont (Rauterberg 1995). Mit Inkongruität ist die Diskrepanz zwischen informationsverarbeitenden, lernenden Systemen und seiner Umgebung als Nachrichtenquelle gemeint. Dieses Einbeziehen von möglichen menschlicher Handlungen zur Bestimmung des Informationsbegriffs wird auch im

Bereich der Kognitionsforschung gefordert. Der Mensch erfährt seine Umwelt nicht nur durch passive Wahrnehmung, sondern zeichnet sich durch aktives Verhalten beim Suchen, Selegieren und Verarbeiten von Information aus der Umwelt aus (Guski 2000).

Andere Ansätze wie der von Herrmann (2001) wenden sich vollständig vom Kommunikationsmodell nach Shannon und Weaver ab und verfolgen ein Modell, mit dessen Hilfe besonders computervermittelte Kommunikation besser abgebildet werden soll. Geht das Modell von Shannon von einer objektiven Realität aus, die die Basis des Zeichenvorrats und des Kanals bildet, lässt das von Herrmann vorgestellte Modell so viele Realitäten (oder auch Wirklichkeiten) zu, wie es Individuen in der Kommunikationssituation gibt. Lediglich ein Teil des äußeren Kontextes kann als gemeinsam bezeichnet und zur Erleichterung der Kommunikation herangezogen werden. Dem Modell liegt eine Haltung zugrunde, die von einer „informationalen Geschlossenheit lebendiger Systeme“ ausgeht (Herrmann 2001), bei der das menschliche Gehirn sich die Vorstellung über seine Umgebung nach der Maßgabe des eigenen Erkenntnisapparates konstruiert. Im Gegensatz zur Mitteilung einer Nachricht über einen möglicherweise gestörten Kanal geht es in diesem Ansatz darum, den Weg der Idee bzw. des Gedankens eines Kommunikationspartners über seinen „kommunikativen Ausdruck“ zu seinem Gegenüber zu beschreiben. Dieser Ausdruck kann zusätzlich noch aus Mitteilungen bestehen, die das sog. extra-kommunikative Handeln begleiten. Bei dieser impliziten Kommunikation spielen nonverbale Äußerungen wie Gesten, Mimik usw. eine Rolle. Implizite Kommunikation zwischen Menschen kann aber auch durch Artefakte der Umgebung vermittelt werden, also indirekt stattfinden. Diese Art des Modells erscheint sinnvoll, wenn Phänomene erklärt werden sollen, die über die reine Übermittlung von Information hinausgehen und die auf den vielfältigen Ausdrucksmöglichkeiten der menschlichen Kommunikation beruhen. Mögliche Kategorien dieser Art der Kommunikation zwischen Mitarbeitern in Bürogebäuden werden z. B. von Kraut et al. (1990) angegeben und reichen von geplanten und vorbereiteten bis zu opportunistischen und spontanen Unterredungen. Gaver (1991a) unterstützt derartige Formen der Kommunikation in Büroumgebungen mit den Mitteln akustischer Darstellungen (Environmental Audio Reminder, EAR).

2.2 Interaktionsmodelle

Wird das Zusammenwirken von Mensch und Computer in seine funktionalen Bestandteile und Zusammenhänge aufgeteilt, entstehen zwei Systeme, die miteinander kommunizieren und ein ‚Regelkreis‘ bilden. Zwischen den Subsystemen wird eine Schnittstelle definiert. Dieses reduktionistische Modell erfüllt für viele Anwendungen und Betrachtungen seinen Zweck, wenn auch die Vereinfachung in manchen Fällen der

Komplexität von realen Situationen nicht vollständig gerecht wird. Im Gegensatz zum Modell nach Shannon und Weaver wird hier die strukturierte Beziehung zwischen Sender und Empfänger berücksichtigt. Bei der Interaktion findet ein Austausch von Botschaften statt. Deshalb kann zwischen der Rolle des Senders und des Empfängers nicht mehr explizit unterschieden werden. Die Schnittstelle ist die Stelle, an der die Informationskanäle aufeinander treffen und die Regelschleife zwischen den beiden Subsystemen geschlossen wird. Die Schnittstelle (engl. Interface) trennt den Menschen vom Computer und damit auch die Bereiche, in denen Information aufgenommen und abgegeben wird.

„Modalität“ bezeichnet die Art und Weise, in der Ereignisse und Objekte aus der Umwelt vom Menschen wahrgenommen werden (Perzeption) und Kontrolle über Objekte ausgeübt wird. Im Gegensatz dazu soll der Begriff Medium die Art bezeichnen, in der Zustände des Computers dem Menschen angezeigt werden, wie z. B. durch optische Darstellungen. Darüber hinaus kann aber u.a. noch von einem Repräsentationsmedium gesprochen werden (Steinmetz 2000), mit dem auf unterschiedliche Formen der computerinternen Darstellung von Information hingewiesen wird. Der Mensch ist bei der Interaktion in zwei unterschiedliche Prozesse involviert, die ihn mit dem Computer in Verbindung bringen:

- die Perzeption von Reizen, d. h. die Aufnahme z. B. eines optischen oder akustischen Reizes der vom Computer ausgesendet wird durch ein Sinnesorgan, und
- die Kontrolle oder Steuerung, d. h. eine intentionale Manipulation von Eingabegeräten des Computers.

Der Computer besitzt Komponenten, die als komplementär zur Perzeption und Steuerung des Menschen bezeichnet werden können. Sensorische Vorrichtungen bzw. Eingabegeräte des Computers sind z. B. die Tastatur, aber auch ein Mikrofon, die dem Menschen die Möglichkeit bieten, gezielte Kontrolle auf den Computer auszuüben. Die inneren Zustände des Computers können z. B. mittels eines optischen Displays dem Menschen zur Wahrnehmung angeboten werden, womit der Kreis geschlossen ist (Feedback). Neben dieser Feedbackschleife kann es beim Menschen zusätzlich zu einem intrinsischen Feedback kommen, z. B. bei der Auge-Hand-Koordination. Hierbei wird über das Auge die Position der Hand wahrgenommen (Rezeption), während die Hand eine Tätigkeit ausführt (Kontrolle). Es kann im Regelkreis zu Zuständen kommen, die vom Menschen nicht bewusst wahrgenommen werden und sich einer gezielten Verarbeitung entziehen, wie z. B. bei Reflexen oder Wahrnehmungen unterhalb der Bewusstseinsschwelle (Rasmussen 1986). Es wird deutlich, dass unter der Berücksichtigung der verschiedenen Eigenschaften der menschlichen Wahrnehmung

und der Einbeziehung der Ziele der Interaktion das oben beschriebene Modell an Grenzen stößt und entsprechend erweitert werden muss.

Mensch wie auch Computer werden im Interaktionsmodell als Systeme dargestellt, die hinsichtlich ihrer Eingangssignale eine Verarbeitung aufweisen, die sich wiederum auf das Ausgangssignal auswirken kann. Der Computer soll hier als System verstanden werden, welches zwar komplexe interne Zustände besitzt, das aber zu jedem Zeitpunkt determiniert ist. Die interne Verarbeitung im Menschen wird z. B. durch das Modell einer sequenziellen Verarbeitung in den perzeptuellen, kognitiven und motorischen Bereichen angenähert (Card et al. 1983). Mit dem perzeptuellen Modell wird beschrieben, wie ein sensorischer Reiz gewandelt wird, damit dann ein ‚Objekt‘ der Wahrnehmung zur kognitiven Verarbeitung weitergereicht werden kann. Diese Verarbeitung kann wiederum dazu führen, dass dann durch motorische Effektoren (Muskeln etc.) Handlungen erfolgen. Darüber hinaus treten aber auch innere Zustände auf, die unabhängig von äußeren Einflüssen entstehen können.

Bei einer sequenziellen Darstellung des Modells erscheint die Verarbeitung implizit mit der Struktur vorgegeben zu sein. Verarbeitung kann aber auch explizit durch Ziele und Anschauungen gesteuert sein (Schomaker et al. 1995). Dem Benutzer (und teilweise auch dem Computer) fällt dann die Aufgabe zu Prozesse zu teilen und zu delegieren. Der Ablauf der Prozesse und die entstandenen Teilergebnisse müssen explizit koordiniert und integriert werden. Dieser Aspekt fügt dem Modell der Interaktion eine Intention hinzu: Unter der Einbeziehung eines Zielzustandes können Modelle aufgestellt werden, mit denen das intentionale Verhalten des Menschen näher beschrieben werden kann. Die Modelle dienen auch dazu, die Kommunikation über Systeme bei der wissenschaftlichen Analyse zu erleichtern. Es werden begriffliche Grundlagen geschaffen, die für alle am Prozess der Gestaltung der Interaktion beteiligten Personengruppen als Leitfaden dienen. Mithilfe eines derartigen Modells kann über verschiedene Konzepte diskutiert und Anforderungen können definiert werden (Herrmann 1986). Dazu zählen z. B. das IFIP-Modell (Dzida 1983), der Model Human Prozessor (Card et al. 1983) oder das Schichtenmodell (Norman 1986). Das Interaktionsmodell nach Rasmussen (1986) ist im Kontext dieser Arbeit von besonderer Bedeutung, da es mehrere Modalitäten, Intentionen sowie Handlungssteuerung berücksichtigt. Im Zentrum des Modells steht ein dynamisches Weltmodell des Benutzers, das sich über die z. T. unbewusste Wahrnehmung der Umwelt bilden und verändern kann.

2.2.1 Wahrnehmung

Im allgemeinen Modell der Interaktion bezeichnet die Wahrnehmung (Perzeption) den bewussten psychischen Vorgang, der auf die Auffassung der unmittelbaren Wirklichkeit gerichtet ist. Veränderungen in der physikalischen Umwelt werden als Reize durch die Sinnesorgane aufgenommen, weiterverarbeitet und vom Menschen bewusst interpretiert. Die Disziplin der Psychophysik beschäftigt sich mit der Beschreibung und Erforschung der Wahrnehmung von physikalischen Sinnesreizungen und wurde von Gustave Theodor Fechner (1801 - 1887) begründet. Während die physikalische Größe, die die Sinne erregt, messbar ist, kann der gegenständliche Inhalt des Vorgangs der Wahrnehmung beim Menschen nicht eindeutig bestimmt werden. Durch Experimente mit Versuchspersonen können Aussagen und Gesetzmäßigkeiten gefunden werden, die eine quantitative Beschreibung der Beziehung zwischen einer physikalischen und einer psychometrischen Variable ermöglichen. Die Selbstbeobachtung (Introspektion) einer Versuchsperson ist dabei ein wichtiger Bestandteil psychophysikalischer Experimente.

Die menschlichen Sinne lassen sich nach dem britischen Psychologen Sterrington in drei Gruppen unterscheiden (zitiert nach Guski 2000), in denen Reize bzw. Informationen aufgenommen werden über:

- die Umwelt (Exterozeptoren),
- die Lage des Körpers und Zustände der Gelenke und Muskeln (Propriozeptoren),
- den Zustand von Organen (Interozeptoren).

Die Exterozeptoren werden weiter in Kontakt- und Distanzrezeptoren unterschieden, wobei üblicherweise die letzteren als Sinne (Gesichts-, Gehör- und Geruchssinn) verstanden werden.

Im Kontext der Mensch-Computer Interaktion sind insbesondere der visuelle, der auditive und der taktile Sinn der Wahrnehmung zu nennen, während die übrigen Sinne von geringerer Bedeutung sind. Den Begriffen der Wahrnehmung mit verschiedenen Sinnen, sollen die Begriffe des physikalischen Mediums wie folgt gegenüberstehen: visuelle, optische oder taktile Wahrnehmung durch optische, akustische oder haptische Darstellungen.

Die Unterscheidung in einen physikalischen (Medium) und einen psychologischen (Modalität) Bereich soll berücksichtigen, dass z. B. ein akustisches Schallereignis durch die Wahrnehmung zu einem auditiven Ereignis beim Hörer führt. Diese beiden Ereignisse sind voneinander zu unterscheiden, denn von der Korrektheit der Wahrnehmung (*Veridikalität*) kann nicht ausgegangen werden. Bei unterschiedlichen

Personen können dieselben Dinge oder Ereignisse zu verschiedenen Erlebnissen führen. Dieselbe Person kann sogar zu verschiedenen Zeitpunkten gleiche Ereignisse unterschiedlich wahrnehmen (Guski 2000). Es kann somit zu Widersprüchen zwischen der physikalischen und der wahrgenommenen Eigenschaft kommen (Sinnes-täuschungen, Unzuverlässigkeit subjektiver Wahrnehmung). Die Widersprüchlichkeiten zwischen dem Stimulus und dem dadurch resultierenden wahrgenommenen Objekt oder Ereignis wurden im visuellen Bereich aufgezeigt und demonstriert (z. B. in Form von endlosen Treppenstufen). Ein Beispiel aus dem auditiven Bereich sind die sog. *Sheppard-Risset Töne*, die den Eindruck einer sich endlos in Richtung höherer Frequenzen hin bewegenden Tonfolge bei Hörern bewirken (nach Deutsch 1995). Dieser Effekt kann z. B. für die Mensch-Computer Interaktion verwendet werden, um das ‚Scollen‘ am Bildschirm oder langanhaltende Systemprozesse akustische darzustellen (Beaudouin-Lafon und Conversy 1996).

Die Arbeiten des Psychologen James J. Gibson (1986) zur sog. *Ecological Psychology* erweiterten die Ansätze der Gestaltpsychologie, indem sie beispielsweise die Handlungsmöglichkeiten und -notwendigkeiten der Lebewesen bei der Wahrnehmung ihrer Umwelt berücksichtigten (Guski 2000). Die Übersetzung ins Deutsche mit Umweltpsychologie oder ökologische Psychologie birgt die Gefahr falscher Interpretationen in Richtung von Umweltschutz. Im Rahmen dieser Arbeit soll der Begriff ökologische Psychologie verwendet werden. Es wird davon ausgegangen, dass Reize aus der Umwelt Informationsquellen sind, die vom Individuum nach einer sensorischen Verarbeitung wahrgenommen werden. Das Individuum reagiert mit einer Erwartung auf dieses Angebot und überprüft durch Handlung, ob die Information gültig ist und ob sie mit dem vorhandenen inneren Bild bzw. Modell der realen Umwelt übereinstimmt. Der Ansatz zeichnet sich besonders dadurch aus, dass der Prozess der Interaktion aus einem ‚Suchen‘ des Individuums nach Information besteht (siehe auch Streitz 1985). Es werden keine Repräsentanten realer Gegenstände in Form von ‚mentalenen Objekten‘ gebildet, sondern die ‚ökologische‘ Information muss nur aufgenommen zu werden. Der Mensch wird nicht als passives, Information konsumierendes Subjekt verstanden, sondern interagiert aktiv mit der Umwelt: Wahrnehmung dient der Steuerung zielgerichteter Handlungen (Guski 2000). Auf die Bedeutung der ökologischen Psychologie im Zusammenhang mit akustischer Wahrnehmung wird in Abschnitt 3.2 näher eingegangen.

Die zwei Seiten der Information im erweiterten Sinne (Erstmaligkeit und Bestätigung nach Weizsäcker (1974), siehe Abschnitt 2.1) deuten sich in Gibsons Auffassung von Information bei Interaktion des Menschen mit seiner Umwelt an. Der Mensch hat Erwartungen an die reale Umwelt aufgrund seiner bisherigen Erfahrungen und aktuellen

Wahrnehmung. Diese Erwartungen sind dann besonders konkret, wenn aus vorher gemachter Erfahrung angenommen werden kann, dass es sich um lebensdienliche Informationen handelt (Resonanz) oder wenn die Angebote, die den Objekten der Umwelt innewohnen, zur Aufnahme auffordern. Neue Informationen, die keiner Erwartung entsprechen, müssen mit Aufmerksamkeit bedacht und deren Bedeutung für Handlungsmöglichkeiten erlernt werden.

Gibson unterscheidet die Wahrnehmungswelt in Substanzen und Oberflächen. Die Art, wie die Oberflächen gestaltet sind, hat Einfluss darauf, welchen Charakter die dahinter angeordnete Substanz vermittelt und welche Wahrnehmung über die Oberfläche angeboten wird (z. B. vermitteln raue und unebene Oberflächen Unruhe und Unsicherheit usw.). Diese Eigenschaft von Oberflächen nennt Gibson *Affordance*. Gibson betont, dass seine Wortschöpfung die komplementären Beziehungen der Lebewesen zur Umwelt hervorheben soll. Diese Beziehungen können jedoch nicht gemessen werden wie physikalische Eigenschaften (Gibson 1986). Guski schlägt für die Übersetzung des Begriffs *Affordance* ins Deutsche den Ausdruck *Passung* zwischen Umweltgegebenheiten und Organismen vor (Guski 2000). Er hebt hervor, dass es sich bei den *Affordances* nicht um vorübergehende und bedürfnisabhängige Angebote handelt. Primär wird die *Affordanz/ Passung* eines Gegenstands vom Menschen durch seine wahrnehmbaren Eigenschaften erfahrbare. Norman (1988) und Gaver (1991b) greifen dieses Konzept insbesondere für die Gestaltung der Mensch-Computer Interaktion auf.

Wichtig ist jedoch, dass die Informationen über die Passungen im Stimulus selbst enthalten sind und durch geeignete Wahrnehmungssysteme entdeckt werden können. Als Beispiel hierzu führt Gaver (1991b) den Türkbeschlag an, deren vertikale oder horizontale Anordnung dem Benutzer nahe legt, ob er die Tür ziehen muss oder ausdrücken kann. In der Wahrnehmung der Türklinke ist also auch die Information, in welche Richtung diese zu öffnen ist, schon enthalten (Gaver 1991b)¹. Dieser Auffassung der Wahrnehmung als Aufnahme von Objektinformation und den damit verbundenen Handlungsmöglichkeiten kann auch helfen, akustische Repräsentationen in geeigneter Weise für die Mensch-Computer Interaktion zu bilden.

¹ Elektrische Schiebetüren, die sich bei der Annäherung von Personen automatisch öffnen, sind um eine derartige *Affordanz* beraubt. Die Bewegung einer Person in Richtung auf eine Tür impliziert nun deren Öffnung.

2.2.2 Multimodale Interaktion

Bei der Interaktion zwischen Mensch und Computer sind in realen Situationen mehrere Medien und Modalitäten involviert. Diese können sowohl sequentiell als auch simultan verwendet werden und zu sehr komplexen Abläufen führen. Nach Schomaker et al. (1995) sollte eine Interaktion dann als multimodal bezeichnet werden, wenn sie die folgende Eigenschaft erfüllt:

„ ... multimodality is restricted to those interactions which comprise more than one modality on either the input (i. e., perception) or the output (i. e., control) side of the loop *and* the use of more than one device on either side“ (Schomaker et al. 1995, Seite 6).

Diese Definition schließt dann Formen der Interaktion aus, bei denen eine der beiden Verbindungen zwischen Mensch und Computer nur über *ein* Medium oder *eine* Modalität erfolgt oder nur ein Ein- oder Ausgabegerät am Computer verwendet wird. Eine Situation, bei der ein Mensch ein Gerät nur über *einen* Schalter bedient und auditives wie auch taktilen Feedback erhält, ist dann keine multimodale Interaktion. Im geschlossenen Regelkreis bedeutet dies einen Engpass für die Regelschleife bei komplexen Interaktionen.

Da diese quantitative Beschreibung auf viele Situationen der Interaktion zutrifft, kann eine weitere Forderung an ein multimodales System gestellt werden, um Interaktionen wie z. B. Autofahren auszuschließen. Nach Nigay und Coutaz (1993) ist für ein multimodales System entscheidend, dass Informationen nicht nur über mehrere Kanäle zum Benutzer geführt werden. Aus der Information über die Interaktion soll auch eine gewisse Bedeutung für den Benutzer abgeleitet werden können. Diese Transformation kann z. B. bewirken, dass aus den Daten der Benutzerinteraktion im Computer symbolische Repräsentationen erzeugt werden, die dann weiterverarbeitet werden können. Ein System mit einer Spracheingabe, bei der das Signal lediglich aufgezeichnet wird, kein multimodales System, dagegen kann ein System, dass aus Sprachsignalen Befehle ableitet, als multimodal bezeichnet werden. Ist also eine Transformation zwischen Information und Bedeutung bzw. ein Übergang von syntaktischer zu semantischer Verarbeitung ein Teil der Benutzerschnittstelle, weist dies auf ein multimodales System hin.

Der Ansatz multimodaler Interaktion zeichnet sich dadurch aus, dass die Verwendung mehrerer verschiedener Kanäle im Sinne der Kommunikationstheorie nach Shannon und Weaver eine Vergrößerung der Übertragungsbandbreite bedeutet, die folgende Vorteile mit sich bringen kann (Robertson 1998):

- A) Größere Bandbreite bei der Verarbeitung im Regelkreis der Interaktion,
- B) manche Modalitäten eignen sich besonders für gewisse Aufgaben (in Abhängigkeit von Person und Zeit),
- C) die Anzahl der Interaktionsfehler können durch Redundanz verringert werden und
- D) die multimodale Interaktion ist natürlicher und involviert den Menschen stärker bei der Interaktion.

Die Bezeichnung Multimedia soll hier komplementär zu Multimodalität verstanden werden, wobei der Schwerpunkt von Multimedia in der Technologie liegt, mit der über verschiedene Kanäle Information in und aus dem Computer geführt wird. Nach der oben genannten Definition von multimodaler Interaktion enthält diese auch immer die Verwendung mehrerer Medien. Die Definition eines Multimediasystems nach Steinmetz lautet:

„Ein Multimediasystem ist durch die rechnergesteuerte, integrierte Erzeugung, Manipulation, Darstellung, Speicherung und Kommunikation von unabhängigen Informationen gekennzeichnet, die in mindestens einem kontinuierlichen (zeitabhängigen) und einem diskreten (zeitunabhängigen) Medium kodiert sind“ (Steinmetz 2000, Seite 13).

In diesem Sinne ist der ‚Multimedia PC‘ die Bedingung der Möglichkeit von multimodaler Interaktion zwischen Mensch und Computer. Zur Beschreibung und Einordnung verschiedener Ansätze zur Gestaltung einer Benutzungsschnittstelle werden in englischsprachigen Publikationen oft eine Vielzahl von Begriffen verwendet, die suggerieren, dass völlig verschiedene Formen der Interaktion vorliegen. Begriffe, wie *perceptive user interface* oder *perceptual user interface* (Turk und Robertson 2000) führen aber eher zu Irritationen als dass sie helfen, die Kommunikation verschiedener Fachgruppen über das Thema zu fördern. Im Rahmen dieser Arbeit soll der Begriff der multimodalen Benutzungsschnittstelle bevorzugt werden.

Eine mögliche Klassifizierung von Systemen mit multimodaler Interaktion wird von Nigay und Coutaz (1993) vorgeschlagen. Die Dimensionen dieser Klassifizierung bzw. des ‚Designspace‘ sind die Fusion der Daten der Ein- und Ausgabegeräte und die zeitliche Verwendung bzw. Anordnung von verschiedenen Modalitäten bei einer typischen Interaktion. Damit können vier Bereiche von wechselnder, synergetischer, exklusiver und gleichzeitiger Interaktion definiert werden (siehe Abb. 3).

		Benutzung der Modalitäten	
		sequentiell	parallel
Fusion der Daten	kombiniert	wechselnd	synergetisch
	unabhängig	exklusiv	gleichzeitig

Abb. 3: Klassifikation multimodaler Schnittstellen nach Nigay und Coutaz (1993).

Die Einordnung nach dem Grad der Fusion von Daten einer Interaktion erfolgt danach, ob diese Daten über die entsprechenden Kanäle nur Teil einer übergeordneten Information (Kombination) sind oder ob sie unabhängig voneinander einen Sinn bei der Interaktion machen. In der anderen Dimension, ‚Benutzung der Modalitäten‘ wird eine Interaktion danach eingeordnet, wie die Verwendung der beteiligten Modalitäten temporal angelegt ist. Dieser zeitliche Aspekt ist sicherlich entscheidend, um z. B. von einer Erhöhung der Effizienz durch eine multimodale Interaktion sprechen zu können. Der Vorteile multimodaler Interaktion nach Robertson (1998) lassen sich in der Klassifikation von Nigay und Coutaz (1993) wiederfinden: z. B. korrespondiert die größere Bandbreite (Vorteil A) mit der parallelen Nutzung der Modalitäten.

2.3 Die Rolle des Computers

Der Computer hat eine besondere Stellung unter den Maschinen, da er verschiedene Maschinen simulieren und zum anderen mit ihm wiederum auch Maschinen hergestellt werden können. Bei der Verwendung eines Computers wird im übertragenen Sinne auch von Werkzeug oder Maschine gesprochen, obwohl der Computer hauptsächlich symbolische Repräsentationen verarbeitet. Häufig wird vom Computer auch als ‚Universalmaschine‘ gesprochen. Zusätzlich fallen mit dem Computer die drei Funktionen der Speicherung, Übertragung und Verarbeitung von Daten zusammen.

Je nach Verwendungskontext nimmt der Computer eine bestimmte Rolle für den Benutzer ein. Um z. B. eine Aufgabe zu bewältigen, wird vom Computer als ‚Werkzeug‘ gesprochen. Dies suggeriert eine direkte Bearbeitung des Problems mithilfe des Computers, während die Vorstellung des Computers als Maschine oder Rechner schon mehr auf Effekte der Automatisierung von Arbeitsprozessen abzielt. Der Computer als Werkzeug verhält sich weitgehend passiv und die Interaktion mit ihm ist

auf ein Ziel hin gerichtet. In anderen Situationen des Umgangs mit Computern kann dieser die Rolle eines virtuellen ‚Dialogpartners‘ einnehmen, mit dem kommuniziert werden kann. Die Begriffe Dialog und Partner sind hier als Metapher zu verstehen, die aber nicht für alle Situationen der Aufgabenbewältigung geeignet oder anzustreben sind. Im Allgemeinen sind beide Rollen in Systemrealisierungen nicht scharf voneinander getrennt und können koexistieren. Während z. B. die Planung und Organisation von Aktivitäten und Terminen mithilfe eines ‚Dialogpartners‘ erfolgen kann, ist bei der Erstellung von Texten die Verwendung eines ‚Werkzeugs‘ zur Textverarbeitung sinnvoll (Schomaker et al. 1995).

In weiteren Situationen, in denen der Computer die Aufgabe hat, die Kommunikation zwischen Menschen zu ermöglichen oder auch zu verwalten, kann er die Rolle eines ‚Mediums‘ einnehmen. Der Computer kann zusätzlich auch als ein Medium zwischen einem Systemdesigner und dem Benutzer aufgefasst werden (Streitz 1985). Der Systemdesigner induziert beim Benutzer ein gewisses Modell, das dem Benutzer bei der Interaktion mit dem System helfen soll. In Analogie zu einer Sprecher-Hörer Situation geht es dann nicht nur darum, sich in verständlicher Form mitzuteilen (mittels einer syntaktisch eindeutigen Interaktionssprache), sondern auch, dem Benutzer bei der Rekonstruktion des intendierten Systems zu helfen und ihn anzuleiten (Herrmann 1986).

Damit ergeben sich drei Rollen, die der Computer bei der Mensch-Computer Interaktion einnehmen kann, die auch als globale Modelle bezeichnet werden (Streitz 1985):

- Werkzeug,
- Dialogpartner und
- Medium.

Innerhalb dieser globalen Modelle können auch noch spezifische Modelle für die Mensch-Computer Interaktion unterschieden werden. Dabei stellen diese Modelle die verschiedenen Repräsentationen dar, die sich spezielle Personengruppen vom System machen, wenn sie das System entwerfen, das System realisieren oder das System benutzen. Diese Personengruppen können wie folgt unterschieden werden nach (Streitz 1985): Psychologen und Ergonomen, Designern und Ingenieuren sowie Benutzern. Diese drei Personengruppen haben ein spezifisches Modell vom System, das zu verschiedenen Zeiten entsteht und unterschiedliche Zeiträume der Gültigkeit hat. Eine wichtige Unterscheidung in diesem Zusammenhang ist die zwischen mentalen und konzeptuellen Modellen. Dem Benutzer wird durch Handbücher, Betriebsanleitung und Systemoberfläche ein konzeptuelles Modell angeboten, das z. B. Ergonomen und Designer erarbeitet haben. Der Benutzer antwortet auf dieses Angebot mit der

Entwicklung eines mentalen Modells vom System. Das Verhalten des Systems auf Eingriffe durch den Benutzer formt schrittweise dessen Vorstellung der inneren Struktur und der funktionalen Zusammenhänge des Systems. Diese Entwicklung ist eingebettet in die Vorerfahrung des Benutzers mit anderen Systemen, interferiert also mit bereits etablierten mentalen Modellen oder baut auf solchen auf. Im Zusammenhang der Mensch-Computer Interaktion fordert Norman (1988) größtmögliche „Sichtbarkeit“ (*visibility*) der Funktion und Struktur eines Systems über die dem Benutzer verfügbare Systemoberfläche. Ist durch den Designprozess das dem System zugrundeliegende konzeptuelle Modell adäquat auf der Systemoberfläche abgebildet, kann sich das mentale Modell auch schnell entwickeln. Kurze Lern- und Orientierungsphasen können dann im Umgang mit dem System erreicht werden (Norman 1986).

2.3.1 Interaktion in virtuellen Umgebungen

Mit dem Begriff virtuelle Umgebungen wird darauf hingewiesen, dass der Mensch bei der Interaktion mit einem Computersystem von der tatsächlichen Umgebung getrennt ist. Der vom Computer erzeugten virtuellen Umgebung liegt ein Modell zugrunde, das von einem Computer z. B. zu einem optischen Signal verarbeitet und dann den Augen als Reiz angeboten wird. Ziel der virtuellen Umgebung ist es dabei, beim Menschen Empfindungen zu erzeugen, die ihm eine akzeptierte Vorstellung einer ‚Modellwelt‘ liefern. Die technischen Randbedingungen zur Herstellung von virtuellen Umgebungen bestehen aus den räumlichen und zeitlichen Auflösungsgrenzen der menschlichen Wahrnehmung. Zum einen ist die Synchronisation der Darbietungen für die verschiedenen Modalitäten entscheidend. Zum anderen ist die Möglichkeit der Interaktion des Benutzers mit dem Modell des Systems wichtig, ohne die eine virtuelle Umgebung eher einer aufwändigen Animation gleichen würde. Nach Schomaker et al. (1995) zielen virtuelle Umgebungen darauf ab, mit einer immersiven audio-optischen Illusion die Realität zu imitieren. Mit multimodalen Systemen hingegen wird versucht, die Verarbeitung bzw. die Interaktion zu unterstützen und ‚Natürlichkeit‘ bei der Interaktion zu erzeugen, wie z. B. mit Sprach- und Gestensteuerungen. Schomaker et al. (1995) verstehen Forschung im Bereich virtueller Umgebungen als Teilbereich der Forschung auf dem Gebiet multimodaler Systeme.

Virtuelle Umgebungen, in die der Benutzer gänzlich eintaucht und dabei mit keinerlei realem Objekt mehr in Kontakt tritt oder dieses zu sehen bekommt, ist nach Milgram und Kishino (1994) nur das eine Ende des sog. Kontinuums der Virtualität (*virtuality continuum*), während das andere Ende reale Umgebungen bilden. Mit dem Begriff „Mixed Reality“, d. h. gemischte oder vermischte Realität, bezeichnet Milgram solche Umgebungen, in denen realen Elemente von virtuelle Elemente überlagert werden bzw.

realen und virtuelle Elemente vermischt sind. Ein frühes typisches Beispiel der *Augmented Reality* (angereicherte Realität) ist z. B. das *DigitalDesk* von Wellner (1993). Bei dieser Arbeit wurden realen Papierdokumenten auf einer Tischoberfläche durch eine Computerprojektion von oben virtuelle Dokumente hinzugefügt. Darüber hinaus wurde die Möglichkeit geschaffen, durch Interaktion mit einem Stift oder dem Finger die virtuellen und realen Dokumente zu manipulieren. Bei einer künstlerischen Installation von Masaki Fujihata, *Beyond Pages* von 1995 (in Schwarz 1997), werden die Seiten eines Buches in vergleichbarer Weise auf eine Tischoberfläche projiziert und können dort ‚umgeblättert‘ werden. Reale Objekte der näheren Umgebung lassen sich auf speziellen Seiten des virtuellen Buches manipulieren: So kann z. B. eine Schreibtischlampe auf dem Tisch über einen virtuellen Schalter im Buch eingeschaltet werden. Bei dieser Arbeit werden auch akustische Signale verwendet, um z. B. die Interaktion beim Umblättern der Seiten des Buches mit entsprechenden Geräuschen anzureichern. Interessant ist bei beiden Ansätzen, dass die Funktionalität des Computers auf die reale Tischoberfläche gebracht wird im Gegensatz zum ‚Zwang‘ der vollständigen Rekonstruktion des ‚Desktops‘ auf dem Computerbildschirm samt Papierkorb und Ordner (Wellner 1993). Weiser (1991) nennt dies auch die verkörperte Virtualität (*embodied virtuality*) und bezeichnet damit Gegenstände, die sich vom virtuellen Desktop auf dem Computerbildschirm auf die reale Tischoberfläche zurückentwickeln und zu physischen Artefakten werden (Weiser 1991).

Milgram entwickelte 1994 eine Taxonomie der *Mixed Reality* für visuelle Displays verschiedener Klassen, um auf Video- und Computerprojektionen basierend Displaytechnologien einzuordnen (z. B. zweidimensionale oder stereoskopische Monitore). Das Kontinuum der Virtualität (siehe Abb. 4) beginnt mit den alltäglichen, realen Umgebungen und endet mit vollständig künstlich generierten, interaktiven Umgebungen. Milgram nimmt nicht nur ausschließlich computergenerierte Darstellungen, sondern auch videobasierte Darstellungsverfahren in die Betrachtung mit auf.

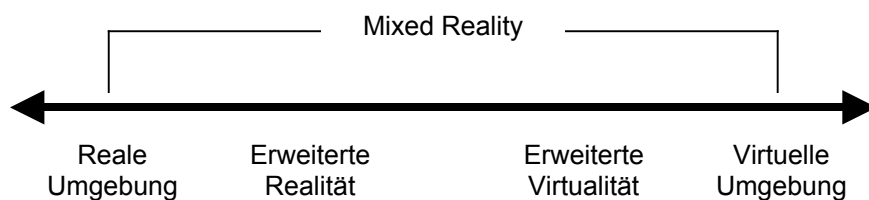


Abb. 4: Einfache Darstellung des Kontinuums der Virtualität (nach Milgram und Kishino 1994).

Er schlägt drei Kriterien vor, nach denen Systeme mit einem Anteil von Video- und/ oder Computerprojektionen klassifiziert werden können:

- den Umfang des Weltwissens: z. B., ob dem Dargestellten etwas hinzugefügt ist, das in der natürlichen Umwelt nicht existiert;
- die Genauigkeit der Wiedergabe, wobei die Qualität vom Drahtgittermodell über Modelle mit Oberflächentexturen bis hin zu fotorealistischen Modellen reicht;
- den Umfang der Metapher der Präsenz²: Der Benutzer kann sich als Beobachter der angebotenen Szene oder in ihr präsent ‚fühlen‘, d. h. er steht außerhalb der Szene oder ist ein Teil von ihr.

Die letzten beiden Kriterien sind nicht ganz unabhängig voneinander. Beide streben auf ein Maximum hin, in dem dann keine Unterscheidung mehr zwischen direkter und generierter Realität besteht. Diese Einstufung beschäftigt sich jedoch hauptsächlich mit den technischen Aspekten der Präsentation, z. B. ob ein *head-mounted display* (HMD) oder stereoskopische Videodarstellungen verwendet werden, und schließt den Anwendungskontext aus.

2.3.2 Ubiquitous Computing

Zwanzig Jahre nach Gründung des *Xerox Palo Alto Research Center* (PARC) verfasste Mark Weiser 1991 einen Artikel mit dem Titel „The Computer for the 21st Century“, in dem er eine Entwicklung in der Verwendung von Computertechnologie antizipierte, die sich vom damaligen Trend der vollständigen Nachbildung von Realität im Computer unterschied. Anstatt mit Datenhandschuh und Stereodisplaybrille den Zugang zu künstlich erzeugten Welten zu schaffen, sollte die Computertechnologie in den Raum integriert werden und somit in ihm ‚verschwinden‘. Mit einem Szenario vom Beginn eines Arbeitstages der Angestellten Sal in Kalifornien illustrierte er anhand einer einfachen Erzählung, was er unter dem Begriff *Ubiquitous Computing* verstand. Damit wurde neben den technischen Aspekten auch gleichzeitig eine mögliche Anwendung im Kontext zukünftiger Büroumgebungen vorgestellt.

Im Zentrum von Weisers Überlegungen steht die Analyse bisheriger Entwicklungen der Computertechnologie und die daraus abgeleitet und aus heutiger Sicht richtige ‚Vorhersage‘ zukünftiger Entwicklungen. In der ‚Blütezeit‘ des Mainframe-Computers teilten sich mehrere Benutzer die Rechnerleistung eines zentralen Großrechners. Mit der

² Hier zu verstehen als Anwesenheit bzw. bewusst wahrgenommene Gegenwärtigkeit in der dargestellten Szene.

Einführung des Personal-Computers (ab ca. 1981), der - wie der Name schon sagt - hauptsächlich von *einer* Person benutzt wird, wurde der Einsatz von Großrechnern immer weiter zurückgedrängt. Weisers Schluss war, dass die Ära des PCs ihrerseits um die Jahrtausendwende dadurch abgelöst werden würde, dass dann Benutzer über mehrere Computer verfügten, die in der Umgebung vorhanden oder eingebettet wären (*Post-PC Ära*; Weiser 1996). Ein Vorbild für eine derartige Entwicklung der Technologie identifiziert Weiser auch in der Elektrifizierung der Industrieproduktion, in der ursprünglich ein Hauptmotor Energie mittels Transmissionsriemen für Hunderte von kleineren Maschinen lieferte. Immer kleinere, günstigere und effizientere elektrische Motoren erlaubten in der ersten Stufe, jede Maschine mit einem eigenem Motor auszustatten und ermöglichten schließlich, dass Maschinen durch eine Vielzahl von Motoren angetrieben wurden. Ein weiterer Aspekt von *Ubiquitous Computing* ist aus Weisers Sicht die vollständige Vernetzung der Computer, so dass der persönliche tragbare Computer mit allen persönlichen Daten unnötig würde. Die dadurch entstehende komplexe Infrastruktur soll eine ähnliche Verfügbarkeit und Unsichtbarkeit besitzen wie die elektrische Stromversorgung im Haus. Zusätzlich fordert Weiser, dass die Computer über Informationen zu ihrer Position und Orientierung im Raum verfügen, um ortsabhängige und kontextbedingte Interaktion für den Menschen zu ermöglichen (z. B. mithilfe von *Active Badge* Technologien; Weiser 1991; Want et al. 1992).

In der Entwicklungsphase zwischen 1988 und 1995 wurden am *Palo Alto Research Center* drei Komponenten für *Ubiquitous Computing* ausgearbeitet *Tab*³, *Pad* und *Boards*, die das Experimentieren mit einem solchen System im Forschungszentrum für eine Gruppe von ca. 40 Personen ermöglichte (Want et al. 1995). Der *Tab* stellt die kleinste Komponente in einer Größenordnung von mehreren Zentimetern dar und kann damit als einer der Vorgänger des heutigen *Personal Digital Assistant* (PDA) bezeichnet werden. Der *Tab* fungiert als persönliches Erkennungszeichen und ermöglicht Anwendungen wie Kalender, E-Mail etc. (Weiser 1991). Für den *Tab* wurde eine auf Infrarot-Technologie basierte Infrastruktur zur Kommunikation mit dem lokalen Netzwerk (local area network, LAN) geschaffen. Die typischen Anwendungen waren das Abrufen und Lesen von E-Mails sowie Wetterinformationen, Dateiverwaltung auf dem Server und das temporäre Zwischenspeichern von Daten zur Verwendung außerhalb der vernetzten Umgebung des Labors.

Want et al. (1995) berichten, dass in einer anderen Installation im *Xerox EuroPARC* Cambridge England, die anfängliche Euphorie über die Personenerkennung mittels

³ *Tab* ist eine Abkürzung für „small tablet computer“ (siehe Want 1995).

Active Badges im Laufe der Zeit jedoch abnahm und sich in Ablehnung verwandelte. Anfänglich war das Lesen der E-Mail auf dem *Tab* sehr populär, doch das Display war dazu eigentlich zu klein und die Bildauflösung zu gering. Lediglich in Besprechungen war dieser Dienst unter den Mitarbeitern deshalb beliebt. Die Komponente *Pad* bestand hingegen aus einem tragbaren Computer mit drahtloser Netzwerkanbindung und einem interaktiven Display in der Größe eines Papierblatts. Das *Board* schließlich war eine elektronische Tafel, die in der Büroumgebung für verschiedene Zweck genutzt wurde (Weiser 1991).

Der Einfluss bzw. die Bedeutung des Konzepts von *Ubiquitous Computing* spiegelt sich in vielen Forschungsbereichen wider. Das Konzept wirft insbesondere für die Mensch-Computer Interaktion auf verschiedenen Ebenen neue Fragestellungen auf, die auf dem Paradigmenwechsel in der Zuordnung von Mensch und Computer basieren (Weiser 1996). Der Computer ist allgegenwärtig und der Benutzer hat von beliebigen Orten aus Zugriff auf seine Daten. Dieses Konzept steht im Gegensatz zur Vision vom ‚Rasenden Stillstand‘ (Virilio 1990), wonach auf die „Mobilisierung der öffentlichen und privaten Transportmittel die Immobilisierung der [Informations]Übertragungen“ folge. Die Informationen würden demnach nur noch zum Menschen transportiert werden und dieser müsste sich nicht mehr bewegen. In vielen aktuellen technischen Forschungsprojekten spielt jedoch die Mobilität von Benutzern komplexer Systeme bei gleichzeitiger allgegenwärtiger Verfügbarkeit von Informationen eine bedeutende Rolle (Project Aura 2001, Project Oxygen 2001). Auch in den besonderen Ansätzen und Projekten, in denen das akustische Medium bei der Mensch-Computer Interaktion verwendet wird (Mynatt et al. 1998, Sawhney und Schmandt 1999), hat Bewegung im Raum einen wichtigen Stellenwert für die Interaktion.

2.4 Akustische Signale bei der Mensch-Computer Interaktion

Die Sonderausgabe der Fachzeitschrift *Human-Computer Interaction*, Volume 4, aus dem Jahre 1989 kann als entscheidender Impuls für die Erforschung akustischer Signale in der Mensch-Computer Interaktion in jüngster Zeit bezeichnet werden. Buxton (1989) thematisiert in seinem Vorwort insbesondere die Rolle der bis dahin vernachlässigten Klasse der *non-speech audio signals* im Kontext der Mensch-Computer Interaktion, Signalen also, die nicht als verbale Kommunikation zwischen Menschen bezeichnet werden können. In dieser Ausgabe der Zeitschrift sind auch zwei Beiträge zu finden, die als grundlegend bezeichnet werden können: Zum einen ist dies William Gavers Artikel über den *SonicFinder* (Gaver 1989), einer Erweiterung der Benutzungsoberfläche des *Apple Macintosh* durch die Anwendung des Konzeptes der *Auditory Icons*. Diese Arbeit beruht auf einem Artikel aus dem Jahre 1986 und auf Gavers Dissertation über

Everyday listening and auditory icons (1988). Zum anderen findet sich in der genannten Sonderausgabe ein Artikel von Blattner et al. (1989) mit dem Titel *Earcons and Icons*, in dem sie den Zusammenhang zwischen visuellen und akustischen Zeichen näher darstellt und mit den sog. *Earcons* eine Gattung von auditiven Zeichen kreiert, die in den folgenden Jahren die Forschung nachhaltig beeinflusste (z. B. Brewster 1994). Drei Jahre nach dieser Sonderausgabe wurde 1992 eine Konferenz von Gregory Kramer ins Leben gerufen, die sich u. a. der speziellen Thematik der akustischen Darstellung widmet und die seitdem fast jährlich abgehalten wird (*International Conference on Auditory Display*, ICAD). Anfang der 80er Jahre begann die Entwicklung und Einführung der Computer mit hochauflösender Grafikdarstellung und Fenstertechnologie. Beispiele derartiger Systeme sind der *Xerox Star* (1981) und dessen Nach- und Weiterbildungen wie der *Apple Lisa* (1983) und der *Apple Macintosh* (1984). Es verbreiteten sich durch diese Rechnergenerationen auch die grafischen Benutzungsoberflächen (*Graphical User Interfaces*, GUI), die einer großen Anzahl von Menschen eine einfache Benutzung von Computern ermöglichten. Als Folge der Bemühungen um menschengerechtere Arbeitsumgebungen mit Computern sind auch die oben beschriebenen Aktivitäten im Bereich akustischer Signale bei der Mensch-Computer Interaktion zu verstehen.

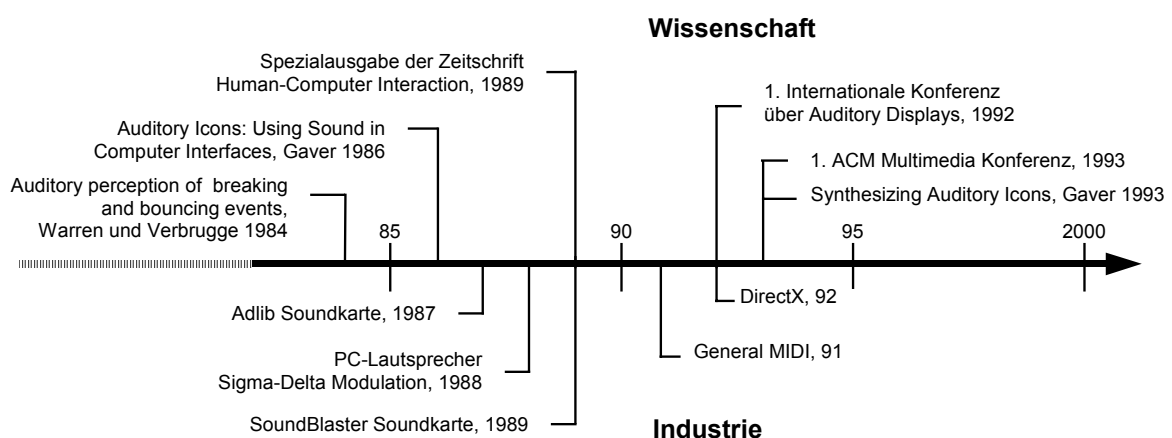


Abb. 5: Zeitleiste der wichtigsten Ergebnisse und Aktivitäten in Wissenschaft und Industrie im Forschungsbereich auditiver Displays im Zeitraum von 1984 bis 1993.

Durch den weiteren Vormarsch von Technologien im Gebiet Multimedia Ende der 80er Jahre gehörte auch eine sog. Soundkarte (z. B. Typ SoundBlaster ab 1989) zur Grundausstattung des Personalcomputers, mit deren Hilfe akustische Signale in den und aus dem Computer geleitet werden konnten (siehe Abb. 5). Neuere Impulse erhält die

Forschung in diesem Bereich durch das Voranschreiten der Miniaturisierung und drahtlosen Vernetzung von Informations- und Kommunikationstechnologie und den damit verbundenen Möglichkeiten bei der Benutzung. Die Ansätze für akustische Darstellungen bei der Mensch-Computer Interaktion beginnen mit Erweiterungen innerhalb von Desktopcomputern und Workstations wie dem SonicFinder von Gaver (1989) und enden mit Projekten, bei denen Technologie eingesetzt wird, die tragbar ist und damit Mobilität für die Anwendung ermöglicht (Bederson 1995, Sawhney 1999, Eckel 2001). War eine Motivation für akustische Darstellungen z. B. die Vergrößerung der Darstellungsmöglichkeiten mithilfe des akustischen Mediums, wie z. B. beim ShareMon (Cohen 1994a) oder bei der sog. Sonification (Kramer 1994), so werden die nachfolgenden Entwicklungen immer mehr durch die Ideen des *Ubiquitous Computing* (Weiser 1991) beeinflusst. Der begehbare Raum wird als Interface benutzt und die akustischen Darstellungen erfolgen in Abhängigkeit der Interaktion des Benutzers im Raum (Mynatt et al. 1998). Selbst wenn mit den akustischen Darstellungen die Eigenschaften des realen Raums übertroffen oder ersetzt werden, bleibt die implizite Interaktion durch die Bewegung im Raum die Basis der Darstellung, wie z. B. im LISTEN-Projekt (Eckel 2001). Aus dieser Tendenz lässt sich zum jetzigen Zeitpunkt antizipieren, dass die Bedeutung akustischer Darstellungen mit der Zunahme und Verbreitung mobiler Informationstechnologie zunimmt.

3 Grundlagen der akustischen Wahrnehmung

Ein wichtiger Baustein für das Verständnis der Bedeutung von akustischen Signalen bei der Mensch-Computer Interaktion bildet die Psychoakustik. Mit ihr werden die Attribute der physikalisch messbaren Schallsignale in Beziehung zu Wahrnehmungseindrücken gesetzt. Einerseits muss derjenige, der für die Gestaltung der Interaktion akustische Signale einsetzt und verwendet, sich darüber im Klaren sein, wie die eingesetzten Signale vom Benutzer wahrgenommen werden und welche Wirkung sie auf diesen haben. Andererseits muss eine Person, die während einer Aktion mit akustischen Signalen konfrontiert wird, selbstverständlich in der Lage sein, die mit ihnen vermittelte Information als solche zu begreifen oder auch zu verstehen. Bei der Verwendung von Geräuschen und Klängen bei der Interaktion können Unterschiede in der Interpretation von akustischen Signalen entstehen. Die Interpretation kann bei Personen unterschiedlich sein oder sich bei derselben Person zu verschiedenen Zeitpunkten ändern. Die Art, wie akustische Darstellungen Informationen vermitteln, kann z. B. auf einem Kontinuum zwischen analoger und symbolischer Darstellung angegeben werden (Kramer 1994). Erkenntnisse aus der Psychoakustik und Betrachtungen z. B. der ökologischen Psychologie spielen daher eine entscheidende Rolle für die Verwendung von Geräuschen und Klängen bei der Interaktion. Die Eigenschaften der akustischen Darstellung können zu denen einer optischen Darstellung komplementär sein (Gaver 1997). Akustische Signale können aber auch visuelle Wahrnehmungen bei der Mensch-Computer Interaktion ergänzen und wie z. B. bei einer multimodalen Interaktion eine kohärente Darstellung und Interaktion bewirken. Shneidermann (1992) weist darauf hin, dass im Bereich der Video- und Computerspielentwicklung die akustischen Effekte schon sehr früh als Mittel verwendet wurden, um dem Spiel einen ‚größeren Realismus‘ zu geben. Begault (1999) berichtet, dass die Verwendung von hochwertigen akustischen Darstellungen bei Computerspielen zur Folge hat, dass die Spieler die optische Darstellungsqualität besser beurteilen als ohne solche akustischen Darstellungen. Oft wird in diesem Zusammenhang auch Jack Foley erwähnt, dessen Name mit der Technik verbunden ist, mit der in den 1950er Jahren Filme im Studio nachvertont wurden (Foley-artist). Foley synchronisierte künstliche Klang- und Geräuscheffekte exakt zu den Szenen, indem er sie beim Betrachten des Films erzeugte und aufnahm (Wyse und Kellock 1999). So wie diese akustischen Signale die Transparenz und Wirkung einer Filmszene erhöhen, können akustische Signale bei der Interaktion mit einem Gerät oder Computer größere Sicherheit und Natürlichkeit bewirken und darüber hinaus Hinweise auf komplexe

Zustände eines Systems geben. Rasmussen (1986) führt dazu das Beispiel an, dass geringfügige Änderungen im Geräusch von Motoren wahrgenommen werden können.

3.1 Psychoakustik

Die Psychoakustik ist ein spezieller Bereich der in Abschnitt 2.2.1 beschriebenen Psychophysik, bei der die Einflüsse von Schallsignalen auf den menschlichen Organismus untersucht werden. Die Schallsignale lösen Prozesse der Empfindung und Wahrnehmung aus und münden in bewusste mentale Hörereignisse. Terhardt (1998) spricht von zwei Klassen von Sinnesempfindungen:

- den *prothetische* Sinnesempfindungen und
- den *metathetische* Sinnesempfindungen.

Die sog. *prothetischen* Sinnesempfindungen vermitteln Quantität bzw. repräsentieren den Aspekt des ‚Wieviel‘ der Stimulation (Intensitätsempfindung). Im Gegensatz dazu geben die sog. *metathetischen* Empfindungen Auskunft über das ‚Was‘ und ‚Wo‘ des Stimulus (Ortsempfindung; Terhardt 1998). Typisches Beispiel der *prothetischen* Empfindung beim Gehör ist die Lautstärkeempfindung, beim Gesichtssinn ist es die Helligkeitsempfindung. *Metathetische* Empfindungen sind z. B. Konturempfindungen von visuellen und auditiven Objekten bzw. der Tonhöhenempfindungen (Kubovy und van Valkenburg 2001). Übliche psychoakustische Experimente bestehen darin, dass Versuchspersonen akustische Signale zugespielt werden und die Personen im Anschluss über ihr Hörerlebnis befragt werden. Die Beschallung erfolgt entweder in *nachhallfreien* Räumen oder mittels Kopfhörer. Dadurch werden die Versuchspersonen definierten akustischen Reizen ausgesetzt und gleichzeitig wird der Einfluss von Störfaktoren auf die Messung verhindert. Durch die Beschreibungen der Hörereignisse durch die Versuchspersonen wird es möglich, Aussagen über quantitativen Beziehungen zwischen Stimulus und Empfindung zu machen. Die Hörsituationen sind abstrakt und unnatürlich wie auch die meisten akustischen Signale, die bei den Versuchen verwendet werden, reine Sinustöne oder Rauschsignale sind (Guski 2000).

3.1.1 Komponenten der auditiven Wahrnehmung

Um in experimentellen Versuchsreihen einen Zusammenhang zwischen physikalischem Reiz und der Empfindungsgröße der Menschen aufstellen zu können, wird der Reiz in grundlegende Komponenten aufgeteilt und im Versuchsverlauf variiert. Die Versuchsperson fungiert dabei als Beobachter des eigenen Wahrnehmungsvorgangs, da sich dieser einer direkten Messung von außen entzieht (Blauert 1997, Terhardt 1989). Die Umstände, unter denen ein Reiz aufgenommen wird, bzw. eine Empfindung

beobachtet wird, sind also immer auch individuell und hängen zusätzlich noch vom gegenwärtigen Kontext des Hörers ab. Elementare Komponenten des akustischen Reizes korrespondieren mit psychometrischen Komponenten wie der Hörwahrnehmung von Lautstärke, von zeitlichen Verhältnissen und von Anteilen unterschiedlicher Tonhöhen.

In diesen drei Bereichen wird jeweils eine Relation zwischen einer physikalischen Größe eines Reizes und einer psychometrischen Größe der Wahrnehmungsempfindung aufgestellt, so dass z. B. ein Zusammenhang zwischen dem Pegel des Schalldrucks eines akustischen Signals am Ohreingang und einer Empfindungsgröße angegeben werden kann. Im Zusammenhang mit einer Interaktion können dann grundlegende Fragen, wie z. B. „ab wann wird ein akustisches Signal wahrgenommen?“ oder „welche Zeitdauer sollte zwischen zwei akustischen Ereignissen mindestens sein?“, beantwortet werden.

Neben dem Schalldruck spielen die zeitlichen Verhältnisse eines akustischen Signals für die auditive Wahrnehmung eine wichtige Rolle. Von einem linearen funktionalen Zusammenhang zwischen den messbaren zeitlichen Verhältnissen eines Signals und seiner mentalen Repräsentation kann nicht ausgegangen werden. Die zeitliche Auflösung in der Wahrnehmung akustischer Signale beträgt ca. 2 ms, um zwei auditorische Ereignisse wahrzunehmen. Unterhalb dieser Schwelle wird lediglich *ein* auditorisches Ereignis wahrgenommen. Nach Pöppel (1997) ist der Zeitbereich bis zur Schwelle der Auflösung geprägt durch die *subjektive Gleichzeitigkeit*. Die Lage der Schwelle ist je nach Modalität verschieden und ist für visuelle Wahrnehmung größer als für auditive Wahrnehmungen. Brewster (1994) argumentiert daher, dass bei der Mensch-Computer Interaktion mit akustischen Darstellungen Informationen schneller vom Menschen aufgenommen werden können als bei optischen Darstellungen. Bis zu einer Zeitdifferenz von ca. 30 ms zwischen zwei akustischen Ereignissen werden diese zwar als einzelne Ereignisse erkannt, jedoch kann ihre Reihenfolge nicht wahrgenommen werden (*unvollkommene Gleichzeitigkeit*). Die Zeitdifferenz von 30 ms wird demnach auch als *Ordnungsschwelle* bezeichnet, da mit größeren Zeitdifferenzen auch die Wahrnehmung der Reihenfolge der akustischen Ereignisse möglich ist. Nach Pöppel (1997) ist diese Ordnungsschwelle unabhängig von der spezifischen Modalität und er schließt daraus, dass es sich um eine höhere Hirnfunktion handeln muss (wie z. B. eine Oszillation der Gehirneuronen). Diese Angaben aus den experimentellen Ergebnissen stimmen mit den speziellen zeitlichen Verhältnissen überein, die bei der Lokalisation von Schallquellen vorherrschen (siehe 3.1.3). Untersuchungen zur Synchronisation von Audio- und Videosignalen zeigen, dass z. B. ein Synchronisationsfehler eines Videos zwischen Bild und Ton, der z. B. durch schlechte

Übertragungsverhältnisse bedingt ist, Testpersonen nicht störte, wenn der zeitliche Versatz innerhalb von -80 ms bis +80 ms bleibt (Steinmetz 2000). Weiterhin konnte ermittelt werden, dass ein Tonsignal, das dem Bildsignal nacheilt, eher toleriert wird als ein dem Tonsignal nacheilendes Bildsignal. Die Begründung ergibt sich aus den unterschiedlichen Laufzeiten des Schalls und des Lichts in natürlichen Umgebungen. Pöppel spricht in diesem Zusammenhang vom *Horizont der Gleichzeitigkeit* (ca. 12 m) innerhalb dessen ein Rezipient *ein* Ereignis wahrnimmt, auch wenn es zu Stimulationen kommt, die zu unterschiedlichen Zeitpunkten beim Rezipienten eintreffen (unvollkommene Gleichzeitigkeit; Pöppel 1997). Jenseits des Horizonts der Gleichzeitigkeit, d. h. der Ordnungsschwelle, führen die zeitlich versetzten Reize zu getrennten Objekten der Wahrnehmung. Für die Verwendung akustischer Darstellungen bei der Mensch-Computer Interaktion sind diese zeitlichen Effekte zu beachten, wenn z. B. mehrere akustische Signale zu *einem* auditorischen Ereignis beim Benutzer führen sollen.

3.1.2 Räumliches Hören

Die Anzahl der Dimensionen eines akustischen Reizsignals kann noch erweitert werden, wenn man zusätzlich den Einfluss des Ortes einer Schallquelle im Raum für die Wahrnehmung berücksichtigt. Das räumliche Hören, d. h. die Identifikation der Relativposition einer Schallquelle zum Hörer, ist eine Wahrnehmungsfähigkeit, die es erlaubt, auch im Dunkeln oder außerhalb des Gesichtsfeldes Ereignisse auditiv wahrzunehmen (Guski 2000).

Aufgrund der Anordnung der Ohren am Kopf kommt es bei der Beschallung mit einer Schallquelle zu Laufzeitunterschieden der Schallsignale an den Ohren in der Größenordnung von wenigen Millisekunden (Interaural-time-delay, ITD) und Schalldruckpegelunterschieden (Interaural-Intensity-Difference, IID) zwischen den Signalen an den Ohren. Durch Beugungs- und Abschattungseffekte des Kopfes und Außenohrs ergeben sich spektrale Verfärbungen der Signale an den Trommelfellen des Hörers. Die Außenohrübertragungsfunktion (englisch: head-related transfer function, HRTF) fasst diese Verzerrung zusammen, die ein Quellensignal in Abhängigkeit des horizontalen und vertikalen Einfallwinkels auf dem Weg zum Trommelfell erfährt. Die Außenohrübertragungsfunktion kann gemessen werden und dann dazu dienen, die Information über die Position der Schallquelle einem Signal ‚aufzuprägen‘. Damit wird es möglich, ein akustisches Signal zu erzeugen, das beim Hörer die Wahrnehmung eines auditiven Ereignisses aus einer bestimmten Richtung bewirkt.

In natürlichen Hörsituationen werden Schallquellen immer in einer Umgebung wahrgenommen, in der die Schallwellen der Quelle nicht nur über den direkten Pfad, sondern auch über weitere Pfade das Ohr erreichen. Diese entstehen durch Reflexionen an den Begrenzungsflächen des Raumes und darin befindlichen Objekten. Die Schallausbreitung in Räumen zwischen einer Quelle und der Position eines Zuhörers können mit den Prinzipien der geometrischen Akustik simuliert werden (Blauert 1997). Mit dem Strahlverfolgungsverfahren werden dazu die linearen Ausbreitungswege von gedachten Schallteilchen beschrieben, wobei beim Auftreffen auf reflektierende Objekte die Reflexionsgesetze angewendet werden. Beim Verfahren der Spiegelschallquellen wird die Schallquelle an allen reflektierenden Oberflächen gespiegelt, so dass eine ‚Wolke‘ von Spiegelschallquellen um den Schallraum herum entsteht. Beide Verfahren sind Näherungsverfahren, haben begrenzte Gültigkeit und werden aus Gründen der effizienten Berechnung bei Schallfeldsimulationen auch kombiniert angewendet. Man kann durch die Simulation die sog. Raumimpulsantwort, d. h. die Übertragungseigenschaften zwischen der Quelle und der Hörerposition bestimmen. Wird noch jeweils die Außenohrübertragungsfunktion für alle Reflexionen berücksichtigt, kann eine sog. binaurale Impulsantwort bestimmt werden, die den Hörer dann auditiv in den simulierten Raum versetzt.

Schallanteile, die mehrmals reflektiert wurden (Reflexionen höherer Ordnung), haben eine geringere Amplitude und treffen deutlich später beim Hörer ein und bilden den sog. diffusen Raumnachhall. Das räumliche Hören in realen Umgebungen fügt dem akustischen Signal neben dem Attribut der Relativposition einer Schallquelle zum Hörer ein weiteres Attribut hinzu, das einen Bezug zum Raum herstellt, in dem der Schall wahrgenommen wird. Der Raum kann damit über die Beschallung indirekt Gegenstand der Wahrnehmung werden (Dutilleux und Müller-Tomfelde 1999). Bei der Verwendung derartiger Attribute von Hörereignissen für die Mensch-Computer Interaktion sollte bedacht werden, dass hierbei insbesondere für interaktive Systeme aufwändige Berechnungen vorgenommen werden müssen (siehe Blauert 1997, Kleiner et al. 1993). Spezielle Ansätze verfolgen das Ziel direkte und recheneffiziente Kontrolle über die perzeptuell relevanten Parameter für die Erzeugung räumlicher Hörereignisse zu liefern (Jot 1999).

3.1.3 Effekte auditiver Wahrnehmung

Im Folgenden werden besondere zeitliche Effekte der auditiven Wahrnehmung näher beschrieben, die für die Mensch-Computer Interaktion relevant sind. Die Zeitverzögerung von zwei Schallquellen, die dasselbe Signal abstrahlen, hat z. B. Einfluss darauf, wie und wie viele Quellen wahrgenommen werden (siehe Abb. 6).

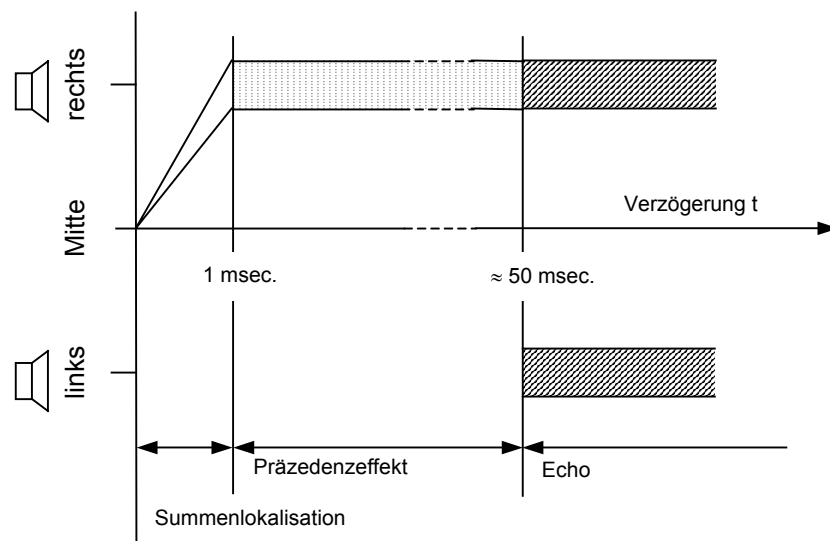


Abb. 6: Darstellung der Lokalisierung einer Schallquelle bei stereophoner Beschallung in Abhängigkeit einer Zeitverzögerung des Audiosignals des linken Lautsprechers, wenn sich die Testperson zwischen den Lautsprechern befindet (nach Blauert 1997).

Bis zu einer Zeitdauer der Lautsprechersignale von 1-2 ms erfolgt die sog. Summenlokalisierung: Wahrgenommen wird dabei eine virtuelle Schallquelle zwischen den Lautsprechern, da die Zeitdifferenz der Signale unterhalb der Auflösung zweier akustischer Signale liegt. Innerhalb eines Zeitfensters von ca. 50 ms lokalisiert der Hörer nun die Quelle mittels des Ohres, an dem die Schallwelle zuerst eintrifft. Der Hörer nimmt nur *eine* Schallquelle wahr und ‚ignoriert‘ das verzögerte Signal. Die Lokalisation der Quellen bleibt selbst dann noch stabil, wenn das verzögerte Signal lauter ist, als das zuerst eingetroffene. Es wird in diesem Zusammenhang vom *Präzedenz-Effekt* gesprochen (Blauert 1997).

Durch die Unterdrückung von Schallanteilen, z. B. ‚früher‘ Reflexionen bei der Wahrnehmung einer Quelle in einer nachhallenden Umgebung ist die ungestörte Ortung der Signalquelle möglich, unabhängig von den konkreten Nachhalleigenschaften des Raums. In diesem Zeitfenster der unvollkommenen Gleichzeitigkeit (Pöppel 1997) wird die Lokalisierung durch das am frühesten eintreffende Signal dominiert. Die Wirksamkeit des Präzedenz-Effekts ist nicht starr an die Ohrsignale gekoppelt, sondern hängt auch vom Gesamtkontext ab, in dem die Schallsignale präsentiert werden (Blauert 1997).

Experimentelle Umgebungen zur Untersuchung der Verhältnisse der Lokalisation von Schallquellen bestehen meist aus einer Anordnung, wie man sie ähnlich aus der

Stereophonie kennt. Der Versuchsperson wird frontal ein Lautsprecher gegenübergestellt und ein weiterer wird seitlich dazu angeordnet, so dass zwischen den Lautsprechern ein Winkel von ca. 30-40 Grad vorhanden ist. Die Zeitverzögerung und die Intensität des Signals des seitlichen Lautsprechers können variiert werden. Die dadurch entstehenden Effekte der Wahrnehmung sind in folgende Bereiche eingeteilt (siehe Abb. 7):

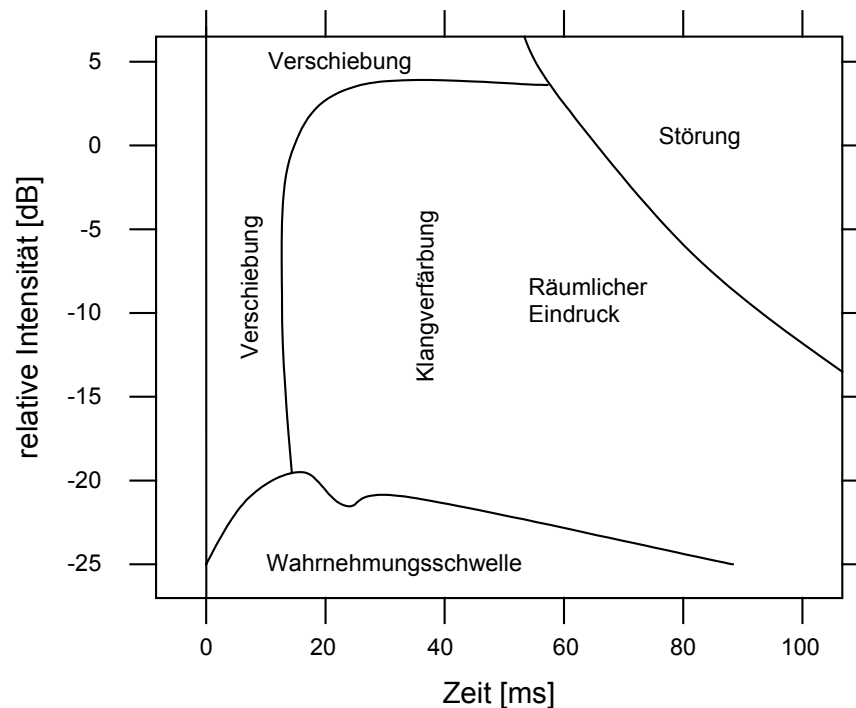


Abb. 7: Effekte der Wahrnehmung bei der Variation der Zeitdifferenz und der relativen Intensitätsunterschiede einer Beschallung durch einen frontalen und einen seitlichen Lautsprecher (nach Kendall 1995a).

- **Wahrnehmungsschwelle:** Hörschwelle in der speziellen Anordnung, bei der der Einfluss des nicht-zentralen Lautsprechersignals ohne Bedeutung für die Wahrnehmung ist.
- **Störung:** Das verzögerte akustische Signal wird als zweite unabhängige Schallquelle wahrgenommen und „stört“ die Wahrnehmung der zentralen Schallquelle (Echo).
- **Verschiebung:** Es wird eine Schallquelle wahrgenommen, deren Position jedoch in der Richtung des lateralen Lautsprechers, also zwischen den beiden Lautsprechern, verortet wird. Dadurch können Schallquellen an Orte zwischen zwei Lautsprechern,

wie in der Stereophonie üblich, platziert werden. Blauert spricht in diesem Zusammenhang von *Summenlokalisation* (Blauert 1997).

- **Räumlicher Eindruck:** Der Zuhörer bekommt den zusätzlichen Eindruck, dass die Schallquelle in einem Raum angeordnet ist. Das akustische Signal gewinnt an „Körper und Klangfülle“ (Kendall 1995a).
- **Klangverfärbung:** Die Schallquelle wird gemäß des zentralen Lautsprechers geortet, jedoch werden die spektralen Eigenschaften wahrnehmbar verändert z. B. durch Kammfiltereffekte.

Der Präzedenz-Effekt erstreckt sich in dieser Aufteilung der Effekte der Wahrnehmung über all die Bereiche, in denen nur die zentrale Schallquelle als Ort des Primärereignisses identifiziert wird, also kein Echo wahrgenommen wird. Weiteren Einfluss auf die Ortung einer Schallquelle bei der Stimulation mit mehreren räumlich verteilten Lautsprechern hat der Grad der *Korrelation* zwischen den Signalen. Signale, die identisch oder zumindest sehr ähnlich sind, verhalten sich nach den oben beschriebenen Regeln. Wird der Grad der Ähnlichkeit z. B. von Rauschsignalen jedoch verringert, nimmt die räumliche Ausdehnung der Hörereignisse zu, bis bei völlig unkorrelierten Signalen wieder einzelne Schallquellen identifiziert werden (Kendall 1995b).

3.2 Ecological Acoustics

In Anlehnung an Gibsons *ecological approach to visual perception* (1986), der speziell auf den Eigenschaften der visuellen Wahrnehmung aufbaut, stellt Gaver in seinem Artikel von 1993c die Frage: „What in the world do we hear?“ und macht den Versuch, auditive Ereigniswahrnehmung aus der Sicht der ökologische Psychologie zu beschreiben. Ein zentraler Ausgangspunkt ist hier wie bei Gibson, dass die Eigenschaften von Objekten oder Ereignissen schon im akustischen Stimulus enthalten sind und vom Menschen direkt aufgenommen werden können. Schon 1984 wiesen Warren und Verbrugge mit Experimenten nach, dass Testpersonen sehr gut zwischen dem Aufprall einer Flasche auf dem Boden mit und ohne Zerschlagen unterscheiden konnten (siehe Abb. 8), selbst wenn z. B. im präsentierten Geräusch der Anteil des initialen Aufpralls entfernt wurde. Dabei spielt vor allem die zeitliche Charakteristik der Energieverteilung im Schallsignal nach dem Aufprall eine wichtige Rolle (Guski 2000).

Bregman (1990) spricht in diesem Zusammenhang von sog. auditory streams (Verkettungen nach Terhardt 1998, Gruppierungen nach Guski 2000). Er weist damit auf einen zeitlichen Informationsfluss hin, mit dessen Hilfe Hörereignisse in eine äquivalente mentale Repräsentation gebracht werden. In Analogie zur Fähigkeit der

menschlichen Wahrnehmung, eine visuelle Szene zu analysieren, entwirft Bregman die Theorie der *Auditory Scene Analysis* (1990).

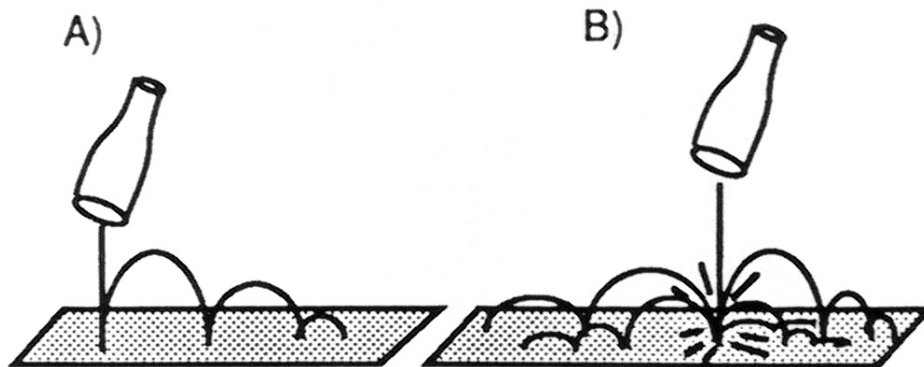


Abb. 8: A) Eine Flasche prallt mehrmals auf den Untergrund. B) Die Flasche zerbricht nach dem Aufprall. Das jeweilige Geräusch ist charakterisiert durch eine zeitliche Abfolge von Einzelgeräuschen (nach Warren und Verbrugge 1984).

Gaver unterscheidet zwei Arten des Zuhörens: das sog. *musical listening* und das *everyday listening*. Während sich das musikalische Zu-hören auf die Komponenten der auditiven Wahrnehmung konzentriert, deren Höreindrücke durch die Gesetze der Psychoakustik beschrieben werden können, sind beim *everyday listening* auch höhere kognitive Prozesse involviert. Nicht nur das Schallsignal, sondern auch die Ursache des Schalls wird zum Bestandteil des Hörerlebnisses (Gaver 1993c). Im Sinne der ökologische Psychologie ist es z. B. in gefährlichen Situationen günstiger, nicht erst die physikalischen Parameter des akustischen Signals wahrzunehmen, sondern sofort und unmittelbar die Ursache zu hören. Das akustische Signal selbst ermöglicht beide Arten des Hörens. Nur die Erfahrung des Hörers und die Situation, in der er sich befindet, entscheiden darüber, auf welchen Gegenstand oder welches Ereignis seine Aufmerksamkeit konzentriert ist.

Die verschiedenen Geräusche, die bei der Interaktion mit Glas entstehen, werden oft für die Situation der Mensch-Computer Interaktion verwendet. Gaver untersuchte die Möglichkeiten und Wirksamkeit der Unterstützung der verteilten Kooperation bei der Simulation einer Flaschenabfüllanlage (Gaver et al. 1991c). Mehrere Mitarbeiter waren räumlich an räumlich getrennte Arbeitsplätze verteilt und hatten eine Abfüllanlage zu bedienen. Verschiedene Geräusche der Flaschen in der Anlage unterstützten die Kooperation und der Klang der ‚Abfüllanlage‘ bei abgestimmten Einstellungen wurde als Ganzes wahrgenommen („Everything sounds alright, doesn’t it?“ aus der

Transkription eines Protokolls, Gaver et al. 1991c, Seite 88). Die Geräusche der Abfüllanlage dienen den Benutzern als Hinweis (Index) auf das Verhalten der Anlage. Hingegen wird z. B. in Programmierumgebungen der Klang von zerbrechendem Glas abgespielt, wenn bei einer Programmübersetzung Fehler auftreten (z.B. NeXtStep, Project Builder). Das Geräusch kann dann als ein auditives Symbol für die Fehlermeldung bezeichnet werden.

3.2.1 Invariante Objektstrukturen

Die physikalischen Gesetze der Schallerzeugung und -ausbreitung ordnen die Umgebung des Menschen, er braucht sie daher nicht kognitiv zu strukturieren (Guski 2000). Zwischen dem Ort der akustischen Quellen und dem Ort der Wahrnehmung existiert ein „Kontinuum strukturierter Energie“, die vom Ort des schallerzeugenden Ereignisses bis zum Gehör des Zuhörers reicht (Gaver 1993c). Aus verschiedenen Beobachtungsrichtungen kann der Mensch sich einen Eindruck von einem Objekt in seiner Umwelt machen. Er bewegt sich damit in einem strukturierten ‚Feld‘ und entdeckt mit jeder neuen Position und Orientierung zu einem Objekt weitere und genauere Informationen. Ein gewisser Teil dieser Informationen ist jedoch unabhängig von der Position und den allgemeinen Parametern der Umgebung und weist damit auf invariante Objektstrukturen hin. Es wird zwischen *strukturellen Invarianzen* und *Transformations-Invarianzen* unterschieden (Warren und Verbrugge 1984; Guski 2000). Als strukturelle Invarianz kann z. B. die Klangfarbe von Musikinstrumenten bezeichnet werden, weil das Instrument vom Hörer identifiziert wird, unabhängig davon, welche Melodie auf dem Instrument gespielt wird. Diese Eigenschaft schallerzeugender Körper oder akustischer Anordnungen existiert in gleicher Weise bei optischen Anordnungen, auch wenn die Gesetze der Ausbreitung der Energie in den ‚Feldern‘ unterschiedlich sind (Gaver 1993c). Der Betrachter entnimmt dem ‚optischen Feld‘ wie dem ‚akustischen Feld‘ aus verschiedenen Perspektiven die strukturellen Relationen, die im Feld konstant bleiben und deshalb als Struktur eines Objektes wahrgenommen werden.

Mit Transformations-Invarianzen sind Objekteigenschaften oder Ereignisse gemeint, die in besonderem Maße durch zeitliche Abläufe offenkundig werden. Die Information über die Art eines Ereignisses wird erst verfügbar, wenn eine Transformation des Objektes selbst erfolgt. Eine Transformations-Invarianz kann demnach das ‚Tanzen‘ einer auf den Boden aufprallenden Flasche sein (siehe Abb. 8) oder die Frequenzverschiebung (Dopplereffekt) des akustischen Signals eines sich mit hoher Geschwindigkeit nähernden Autos. Über die physikalische, invariante Beziehung zwischen der

Frequenzverschiebung und der Änderung der Entfernung gegenüber dem Beobachter wird die Bewegung des Autos als Ereignis wahrgenommen.

Für Warren und Verbrugge (1984) verhalten sich die strukturellen und die Transformations-Invarianzen reziprok zueinander: Während die konstante Struktur des Objekts durch die Veränderung in der Art der Beobachtung offenbar wird, erlauben Transformations-Invarianzen die Vermittlung von Information bei der Beobachtung der Art der Veränderung des Objektes selbst.

3.2.2 Elementare akustische Ereignisse

Gaver (1993c) widmet sich in seiner Beschreibung akustischer Ereignisse insbesondere der Herkunft der Schallwellen aus der Interaktion zwischen Materialien. An der Interaktion am Ort der Schallerzeugung können folgende Objekte oder Elemente beteiligt sein:

- vibrierende Masseschwinger,
- Gase und
- Flüssigkeiten.

Grundlegende Interaktionen dieser Elemente mit- und untereinander bilden die erste Stufe der schallemittierenden Ereignisse (siehe Abb. 9). Besondere zeitliche Muster, wie z. B. beim Aufprall von Objekten auf einen harten Boden entstehen oder die Interaktion mit zusammengesetzten Quellen bilden, komplexe Formen der akustischen Ereignisse. Vier verschiedene Klassen der Interaktion werden z. B. für einen Aufprall besprochen: Ein *grundlegendes Ereignis* stellt ein einzelner Schlag dar, z. B. der eines Hammers auf einen Tisch. Es können dabei elementare Eigenschaften wie die Materialien und Größe der Objekte und der Kraftaufwand etc. vermittelt werden. *Zeitliche Muster* in der Abfolge mehrerer Einschläge können die elementaren Eigenschaften unterstreichen, aber auch besondere Interaktionen signalisieren. *Zusammengesetzte Ereignisse* sind solche, bei denen verschiedene Elementarinteraktionen im zeitlichen Zusammenhang geschehen damit die Komplexität erhöhen und zusätzliche Information vermitteln können. Gaver führt hier die Tätigkeit des Schreibens an, bei der sowohl Aufprallgeräusche des Aufsetzens eines Stiftes auf einem Blatt Papier bzw. der Tischoberfläche als auch das Elementarereignis des Kratzens auftreten. Schließlich stellen *hybride Ereignisse* Situationen dar, in denen verschiedene Materialien am schallerzeugenden Prozess teilhaben, wie z. B. das Tropfen von Wasser auf widerhallende Oberflächen (Gaver 1993c). Alle beschriebenen Formen der Interaktion basieren auf der elementaren Interaktion des Aufpralls.

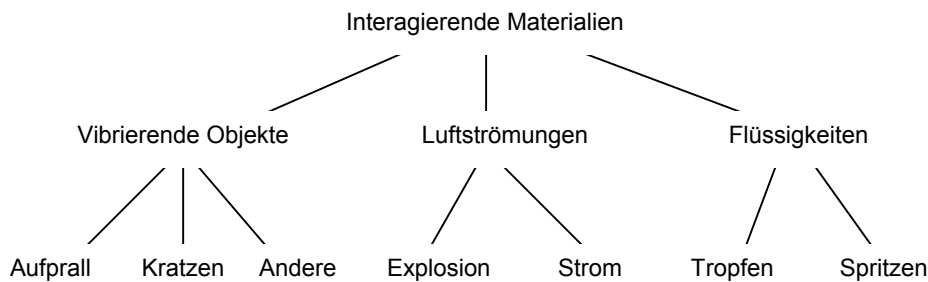


Abb. 9: Hierarchische Struktur einfacher, elementarer akustischer Ereignisse bei der Interaktion von Objekten aus verschiedenen Materialien (nach Gaver 1993c).

Gaver entwirft eine ‚Karte‘ der *everyday sounds*, in der alle oben aufgelisteten Arten von Materialien aufgeführt werden und ordnet ihnen Beispiele der vier verschiedenen Formen der Ereignisse zu. Die Relevanz der Kategorien gewinnt Gaver teilweise aus den Protokollen, in denen Versuchspersonen die Frage nach dem Ereignis gestellt wurde, das sie zuvor zu hören bekamen. Die Dimensionen dieser Darstellung der Geräusche und Klänge des Alltags wie Material, Art des Ereignisses und Komplexität des Gesamtsignals sind voneinander unabhängig. Die ‚Karte‘ ist jedoch nicht vollständig und nicht allgemein anwendbar, da zum einen die Zuordnung stellenweise subjektiv ist und zum anderen nicht alle Ereignisse die gleiche Menge an Information über die Quelle vermitteln (Gaver 1993c). Verwechslungen bei der Identifikation der schallerzeugenden Ereignisse sind weniger zu verstehen als Fehler der Versuchsperson, sondern vielmehr als ein Hinweis auf eine mögliche Gruppierung in einer übergreifenden Kategorie (Warren und Verbrugge 1984).

Für den Bereich der Mensch-Computer Interaktion können akustische Objektstrukturen verwendet werden, um die Interaktion mit Objekten oder zwischen Objekten anzureichern. William Gaver entwickelte den *SonicFinder* (Gaver 1989), eine Erweiterung der Benutzungsoberfläche des *Apple Macintosh*. Dabei wird die Interaktion auf der Benutzungsoberfläche, wie z. B. das Öffnen oder das Verschieben von Objekten der Oberfläche (Dateien, aber auch Fenster), mit aufgezeichneten Geräuschen der alltäglichen Umgebung angereichert. Auf der Grundlage der Karte der *everyday sound* entwickelte Beaudouin-Lafon und Gaver (1994) das ENO-System, das einem Anwender ein Werkzeug zur Verfügung stellt, mit dessen Hilfe komplexe akustische Signale erzeugt werden können, ohne auf der Ebene des Signalflusses explizit ein Geräusch gestalten zu müssen. Neuere Ansätze und Entwicklungen zielen auf die automatische Vertonung von Animationen ab und wurden von Cook (1997) und van Doel (1998) vorgestellt. Diese Ansätze können auch für die Mensch-Computer Interaktion

Verwendung finden, da sie perzeptuell relevante akustische Darstellungen ermöglichen und die Erzeugung keine aufwändigen Berechnungen erfordert.

4 Digitale Signalverarbeitung in Computeranwendungen

Kenntnisse über die technologischen Grundlagen der Signalverarbeitung spielen bei der Realisierung von akustischen Repräsentationen eine wichtige Rolle, gerade vor dem Hintergrund interaktiver Anwendungen. Die zeitlichen Verhältnisse z. B. der Ein- und Ausgabe von Audiosignalen beeinflussen die Qualität der Interaktion stark, denn auch wenn die akustische Repräsentation allen Anforderungen genügt, so führt sie doch zu Irritationen, wenn die ‚Reaktion‘ des Systems auf die Eingabe des Benutzers zu sehr verzögert wird oder sich nur langsam verändert. Damit wird deutlich, dass der Verarbeitung der Audiosignale in Echtzeit eine besondere Bedeutung bei der Mensch-Computer Interaktion zukommt. Die verschiedenen Aspekte der Signalverarbeitung in Computeranwendungen und Verfahren zur Signalerzeugung werden in diesem Kapitel beschrieben.

Mit dem Aufkommen von Multimedia-Technologien Ende der 80er Jahre wurde der Standardcomputer um die Möglichkeit erweitert, akustische Signale abzuspielen und zu erzeugen. Zuvor beschränkte sich die Ausgabe akustischer Signale in einem Personal Computer auf eine einfache Ansteuerung des eingebauten Lautsprechers, mit dem es lediglich möglich war, einen Ton in einer bestimmten Frequenz und Dauer wiederzugeben. Die Rechenleistung dieser Computergeneration erlaubte ihrerseits fast überhaupt keine Verarbeitung von Signalwerten in angemessener Zeit. Digitale Signalverarbeitung war in Echtzeit nur auf dafür konzipierten digitalen Signalprozessoren (DSP) möglich, deren Entwicklung für den Einsatz im Bereich der mobilen Endgeräte der Telekommunikation Mitte der 80er Jahre vorangetrieben wurde.

Wichtige Impulse für die Hardware- und Softwareentwicklung der Signalverarbeitung in Computerumgebungen kamen aus dem Bereich musikalischer Forschung und Komposition, weil dort die Chance gesehen wurde, Klang direkt zu produzieren, mit ihm zu arbeiten und dabei auf ein Musikinstrument im klassischen Sinne zu verzichten. Die Entwicklungen im kommerziellen Bereich beschränkten sich meist auf elektronische Instrumente mit einem festen analogen Verfahren zur Tonerzeugung (Roads 1996). Das Computersystem NeXTstation, das vom Apple Gründer Steven Jobs entwickelt wurde und seit 1989 als Produkt verfügbar war, hatte eine wegweisende Hardwarearchitektur in Hinblick auf die Verarbeitung von Audiosignalen. Neben einer zentralen Recheneinheit und einem mathematischen Koprozessor befand sich direkt auf der Hauptplatine des Computers auch ein digitaler Signalprozessor, der speziell für die Anwendungen der Signalverarbeitung eingesetzt werden konnte.

Die erkennbaren Trends der Signalverarbeitung in Computeranwendungen seit den 80er Jahren können durch zwei Merkmale beschrieben werden:

- die mehrkanaligen Ein- und Ausgabemöglichkeiten von Audiosignalen in Computerumgebungen und
- die Verarbeitung von digitalen Signalen auf dem Hauptprozessor des Computers.

Die Anzahl der Möglichkeiten der Ein- und Ausgabe von Audiosignalen wuchs mit dem Bedarf von Multimedia-Anwendungen auf einem Standardcomputer und beginnt mit der Einführung der SoundBlaster Karte für den Personal Computer 1989. Die seitdem als Standard für Audioein- und -ausgabe geltende Karte wurde immer weiter mit neuen Komponenten wie speziellen Bauteilen zur Erzeugung bzw. Synthese von Audiosignalen ausgestattet. Weiterhin verfügte die Karte über Ein- und Ausgänge zur Kommunikation mit externen Synthesizern (Music Instrument Digital Interface, MIDI). Neuere Anforderungen an die Schnittstelle liegen im Bereich der Mehrkanal-Wiedergabe und der digitalen Ein- und Ausgabe über Formate wie die Anfang der 90er Jahre vorgeschlagene digitale Audio-Schnittstelle AES/ EBU (Audio Engineering Society und European Broadcast Union) und dem Consumer Format SPDIF (Sony Philips Digital InterFace; Zölzer 1997).

Im Forschungsbereich standen weniger die Aspekte der mehrfachen Ein- und Ausgabe als vielmehr die Erzeugung und Manipulation von Audiosignalen im Vordergrund. Dafür wurden oftmals spezielle Hardwaresysteme geschaffen, die mittels eines Computers gesteuert wurden. Repräsentative Beispiele dafür sind die *Samson Box*, benannt nach ihrem Erbauer Peter Samson 1977 und entwickelt am Center for Computer Research in Music and Acoustics (CCRMA). Von Giuseppe di Giugno wurde 1978 am Institut de Recherche et Coordination Acoustique/ Musique (IRCAM) das System *4X* entwickelt (Roads 1996, Smith 1991). Die *Samson Box* entstand aus dem Bedarf heraus, Klang durch Algorithmen mithilfe der Entwicklung spezieller Hardware schnell zu erzeugen, da die Rechenleistung von Computern in den 70er Jahren nicht genügte, um längere Kompositionen durchzuführen. Die *Samson Box* bestand aus 256 diskret aufgebauten digitalen Generatoren, mit denen verschiedene Standardformen der Klangsynthese realisiert werden konnten. Die damit benötigte Berechnungszeit konnte bis um drei Größenordnungen gegenüber der herkömmlichen Art der Berechnung verkürzt werden. Smith spricht von der *Samson Box* aber eher als digitalem Synthesizer denn als Forschungssystem, da keine grundlegenden und neuen Synthesialgorithmen damit entwickelt wurden (Smith 2002).

Das 4X-System hingegen kombinierte einen Standardcomputer als ‚control processor‘ mit einer speziellen Hardware zur Synthese. Mit einem Kontrollprogramm konnte über eine Kommunikationsfunktion auf das Programm zur Klangerzeugung in Echtzeit Einfluss genommen werden. Die Trennung der beiden Funktionseinheiten von Kontrolle und Erzeugung in zwei verschiedene Hardwareumgebungen wurde am IRCAM der 90er Jahre aufgegeben. Mit der IRCAM Music Workstation (IMW) wurde eine Umgebung aus einem RISC-System (reduced instruction set computer) und einem Host-Computer geschaffen. Die Signalverarbeitung und Ereigniskontrolle in Echtzeit wurde dabei auf demselben RISC-System verarbeitet, während die grafische Oberfläche auf dem Host-Computer (einem NeXTCube) dem Benutzer eine einfache Bedienung ermöglichte (Lindemann et al. 1991). Der Entwicklungsschub durch den Erfolg von Anwendungen im Bereich Multimedia und die Entwicklungen aufgrund der Forschungsaktivitäten auf der Suche nach neuen ‚Werkzeugen‘ im Bereich Computermusik streben auf denselben Punkt zu: einen leistungsstarken Standardcomputer mit mehreren Audiosignalein- und -ausgängen.

4.1 Audiosignalverarbeitung in Echtzeit

Eine Verarbeitung in Echtzeit liegt dann vor, wenn die Korrektheit der erbrachten Rechenergebnisse nicht nur von einer fehlerfreien Berechnung abhängt, sondern auch vom Zeitpunkt, an dem die Ergebnisse bereitstehen. Nach der DIN Norm 44300 definiert sich der Echtzeitbetrieb eines Rechnersystems wie folgt:

„Echtzeitbetrieb ist ein Betrieb eines Rechnersystems, bei dem Programme zur Verarbeitung anfallender Daten ständig derart betriebsbereit sind, dass die Verarbeitungsergebnisse innerhalb einer vorgegebenen Zeitspanne verfügbar sind“ (Steinmetz 2000, Seite 232).

Damit sind garantierte Antwortzeiten ein Hauptmerkmal eines Systems im Echtzeitbetrieb und nicht die Geschwindigkeit oder Effizienz der Verarbeitung. Legt man der Verarbeitung das Kommunikationsmodell mit kontinuierlichem Informationsfluss zugrunde kann in Computerumgebungen von Mikrofon, Lautsprecher, Videokamera, aber auch CD-Laufwerk oder Festplatte als Quellen oder Senken gesprochen werden.

Der Computer muss die Übertragung verschiedener Signalströme garantieren, ohne dass es zu auffälligen und störenden Diskontinuitäten oder Artefakten bei der Wiedergabe kommt. Bei heutigen Computersystemen ist diese Eigenschaft bzw. Ressource dem Betriebssystem zugeordnet, das neben seinen Aufgaben der Verwaltung der Ein- und Ausgabegeräte wie Bildschirm, Tastatur und Festplatten auch die Verwaltung der

Rechenleistung des Prozessors (central processing unit, CPU) und des Speicherplatzes bietet. Damit ist es möglich, mehrere Prozesse so hintereinander auf einem Prozessor arbeiten zu lassen, dass von der Seite der Anwendungsprogrammierung und des Benutzers der Eindruck entsteht, dass die Prozesse ‚gleichzeitig‘ bzw. parallel ausgeführt werden. In einem sog. *Multitasking*-Betriebssystem ist ein *Scheduler* dafür verantwortlich, zwischen den verschiedenen Prozessen während der Verarbeitungszeit zu wechseln (üblicherweise geschieht dieser Wechsel in der Größenordnung von wenigen Millisekunden). Weiterhin stellt das *Multitasking*-Betriebssystem Mittel bereit, um über einen gemeinsamen Arbeitsspeicher zwischen den Prozessen Kommunikation und dadurch Synchronisation zu ermöglichen. Prozesse bekommen z. B. eine Priorität zugeordnet, mit deren Hilfe der *Scheduler* über die Reihenfolge der Ausführung der Prozesse entscheidet. Die Prozessorleistung, die ein Prozess beansprucht, ergibt sich aus der Zeitdauer, in welcher der Prozess zur Ausführung kommt, bezogen auf ein bestimmtes Zeitintervall (Steinmetz 2000).

Die besondere Klasse der Echtzeitprozesse zeichnet sich gegenüber den anderen Prozessklassen dadurch aus, dass der entsprechende Prozess in einer festen Taktperiode T_p (siehe Abb. 10) vom *Scheduler* für eine bestimmte Zeit T_e zur Ausführung gebracht wird, um einen gewissen Datendurchsatz zu garantieren. Auch wenn z. B. die Prozessorleistung, die vom Echtzeitprozess P_1 benötigt wird, deutlich kleiner ist als die eines zeitunkritischen Prozesses P_2 muss der *Scheduler* garantieren, dass die Verarbeitung P_1 bis zu einem bestimmten Zeitpunkt beendet ist und gegebenenfalls die Ausführung von P_2 zurückstellen.

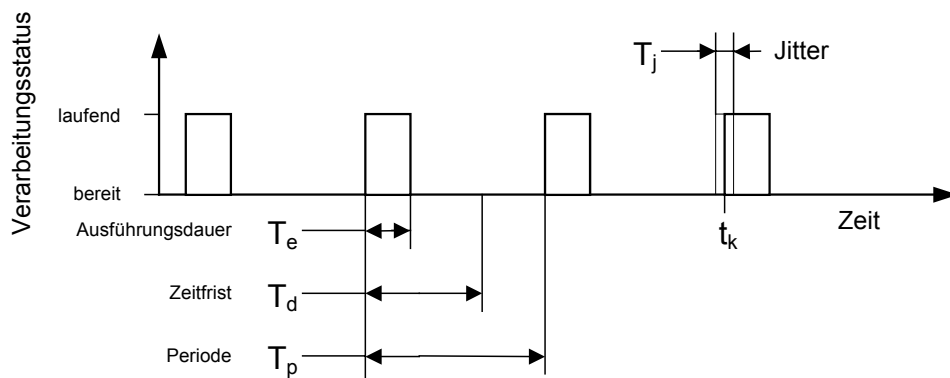


Abb. 10: Zeitliche Verhältnisse eines zeitkritischen Prozesses, der periodisch zur Ausführung gebracht wird. Der Zeitpunkt t_k , an dem der Prozess jeweils neu aktiviert wird, unterliegt systembedingten Schwankungen (Jitter) (nach Steinmetz 2000).

Eine genügend große Rechenleistung ist i. A. nicht ausreichend und kein Garant für eine rechtzeitige und kontinuierliche Verarbeitung von Echtzeitprozessen. Die beanspruchte Prozessorleistung $U(n)$ durch einen periodischen Echtzeitprozess n ergibt sich aus dem Verhältnis von Ausführungszeit und Periode:

$$U(n) = \frac{T_e(n)}{T_p(n)}. \quad (1)$$

Die Gesamtbelastung U_p auf einem Prozessor wird durch die Summe der Leistung aller zur Verarbeitung anstehenden Prozesse N bestimmt, wobei der Bezugszeitraum T_p für alle Prozesse gleich groß sein muss:

$$U_p = \sum_{i=0}^{N-1} U(i) = \sum_{i=0}^{N-1} \frac{T_e(i)}{T_p(i)}. \quad (2)$$

Bei einem einkanaligen Audiodatenstrom mit einer Abtastrate f_s von 44.1 kHz und einer Periode von $T_p = 10$ ms für den zyklischen Prozessaufruf muss z. B. die Verarbeitung von $f_s \cdot T_p = 441$ Abtastwerten durch den Prozess im Zeitraum T_e erfolgen.

4.2 Verzögerungen der Datenströme

Setzt man die zeitlichen Verhältnisse der Verarbeitung wie in Abb. 10 dargestellt in Zusammenhang mit der zeitlichen Struktur des Datenstroms der Eingabe, werden zwei Aspekte deutlich:

- Die Ausgabe der Daten ist um eine konstante Zeit verzögert;
- mehrere Abtastwerte sind zu einem Block zusammengefasst.

Die Verzögerungszeit eines Systems, das periodisch Daten verarbeitet, kann aus der sog. Impulsantwort des Systems gewonnen werden. Die wichtigsten technischen Komponenten, die einen Einfluss auf die Impulsantwort des Systems haben, sind die Umsetzung der analogen Signale in zeitdiskret numerische Daten, der Transport innerhalb eines Rechnersystems und die Verarbeitung durch einen Anwendungsprozess.

4.2.1 Blockverarbeitung

Am Übergang zwischen dem analogen Signal und seiner digitalen Repräsentation befindet sich der Analog-Digital-Umsetzer, der eine Bandbreitenbegrenzung des Signals voraussetzt, um dem Abtasttheorem nach Shannon zu genügen. Diese und weitere Eigenschaften der Umsetzung werden als ideal angenommen, so dass nach dem Umsetzer ein zeitdiskreter und periodischer Datenstrom angenommen wird. Der

Datenstrom besteht aus einer Abfolge von Abtastwerten mit einer Periode von T_s bzw. in einer Abtastfrequenz f_s . Um die Weiterverarbeitung in einem Standardcomputersystem zu ermöglichen, werden die Abtastwerte zu Blöcken bzw. Vektoren von M Abtastwerten zusammengefasst.

Diese Blockbildung von Abtastwerten hat eine signifikante Auswirkung auf die Verzögerung des Gesamtsystems. Eine prinzipielle Verzögerung bei der Echtzeitverarbeitung zwischen Ein- und Ausgang ergibt sich aus der folgenden Betrachtung, wobei keine weitere Zwischenspeicherung durch das Betriebssystem oder die Applikation angenommen wird:

- Es vergeht die Zeit $T_M = M \cdot T_s$, bis ausreichend viele Abtastwerte vorhanden sind, um einen Block für die Weiterverarbeitung zu bilden.
- Dieser Block wird dann in der Applikation in der Zeit T_e weiterverarbeitet; Im günstigsten Fall ist der periodische Aufruf des Echtzeitprozesses $T_p = T_M$ (siehe Abb. 10), so dass ebenfalls nach der Zeit T_M der Block wieder ausgegeben werden kann.
- Aus diesem verarbeiteten Block werden die einzelnen Abtastwerte an den Digital-Analog-Umsetzer weitergereicht.

Die Änderung von serieller zu vektorieller Repräsentation stellt eine zeitliche Entkopplung bzw. eine Zwischenspeicherung von Signalwerten dar. Wird ein ideales Verhalten des Systems in Echtzeit vorausgesetzt, ergibt sich somit durch die vektorielle Verarbeitung eine zeitliche Verzögerung des Ausgangssignals von

$$T_l = 2T_M = 2MT_s = 2 \frac{M}{f_s}. \quad (3)$$

Wie aus Messungen hervorgeht, sind die Werte für die Verzögerungen T_l bei Standardcomputersystemen breit gestreut, wenn verschiedene Konfigurationen der Hardwarekomponenten wie Soundkarte und Betriebssystem verwendet werden (MacMillan et al. 2001, Bencina und Burk 2001). Die Gesamtverzögerung kann von wenigen Millisekunden bis hin zu mehreren hundert Millisekunden reichen. Die entscheidenden Bereiche, die diese spezielle Verzögerung des Datenstroms beeinflussen, sind u. a. an der Schnittstelle zwischen Hardware und Betriebssystem, dem sog. Treiber, und an der Schnittstelle zur Anwendungsprogrammierung zu suchen. Die Effizienz dieser Komponenten, die den Zugriff auf die Datenströme in einer generalisierten Computerumgebung ermöglichen, ist entscheidend für die Größe dieser speziellen Zeitverzögerung bei der Datenverarbeitung. Im Sinne der *Vorhersehbarkeit*

ist es auch von größter Wichtigkeit, dass die Verzögerungszeit des Datenstroms während der Verarbeitung konstant ist. Ein weiterer Anteil der Verzögerung entsteht durch die Anwendung selbst. Werden z. B. Audiosignale gefiltert, addiert sich zu den vorher beschriebenen Zeitspannen die Laufzeit des Signals durch das Filter.

4.2.2 Behandlung von Ereignissen

Neben der kontinuierlichen Verarbeitung und Übertragung von Datenströmen können auch Ereignisse, die z. B. vom Benutzer über die Tastatur oder Maus erzeugt werden, Einfluss auf die Signalverarbeitung haben. Neben der Verzögerungszeit, die zwischen dem Signalein- und -ausgang herrscht, können auch Verzögerungszeiten entstehen, die zwischen dem Eintritt eines Ereignisses und der Auswirkung am Signalausgang oder zwischen dem Signaleingang und einem Ausgabeereignis existieren. Diese Verzögerungen sind wichtig, wenn Aspekte der Mensch-Computer Interaktion im Vordergrund stehen. Eine zu große Zeitverzögerung bzw. Antwortzeit zwischen dem Auslösen des Ereignisses und der ‚Reaktion‘ des Systems erscheint widernatürlich und kann zu Irritationen führen.

Bei der Einwirkung eines Ereignisses auf die Verarbeitung des Signalstroms handelt es sich üblicherweise um Kontrollereignisse, mit denen es dem Benutzer der Anwendung ermöglicht wird, z. B. Signalströme miteinander zu mischen oder einen Parameter der Erzeugung eines Audiostroms zu manipulieren. Die Verarbeitung der Ereignisse erfolgt in der Regel nicht mit Echtzeitprozessen, um das Gesamtsystem mit dem geringen Datenaufkommen der Ereignisse nicht unnötig zu belasten und geschieht üblicherweise im Takt der Signalblockverarbeitung (Puckette 1991a, 1991b). In manchen Anwendungen wie in virtuellen Umgebungen kann es jedoch sinnvoll sein, Kontrollereignisse, wie z. B. die Daten eines Geräts, zur Bestimmung der Position und Orientierung des Kopfes des Benutzers nicht nur als Folge von Ereignissen, sondern auch als Signaldatenstrom mit der Abtastfrequenz eines Audiosignals in die Applikation zu leiten. Es wird dadurch eine bessere Stabilität und Synchronisation bei der Verarbeitung erreicht.

4.3 Erzeugung von Audiosignalen

Für die Erzeugung von Audiosignalen können verschiedene Methoden und Ansätze verwendet werden, die jeweils besondere Ziele verfolgen und sich deutlich in ihrem Verarbeitungsaufwand voneinander unterscheiden. Die Bereiche der Multimedia-Anwendungen und der Computermusik können herangezogen werden, um die Verfahren zur Erzeugung von Audiosignalen zu identifizieren. Während sich eine

solche Erzeugung bei Multimedia-Anwendungen in Computerumgebungen zum größten Teil auf das Abspielen von Audiodateien mit der Kontrolle der Abspiellautstärke in Echtzeit beschränkt (Wyse und Kellock 1999), bieten die Entwicklungen aus der Computermusik eine Vielzahl von Möglichkeiten, komplexe Klänge zu erzeugen. Die Annäherung beider Bereiche kann man folgendermaßen beschreiben: Im Bereich der Entwicklung für Multimedia-Anwendungen wird nach effizienten Methoden der Klangerzeugung gesucht, um den Bedarf an Prozessorleistung möglichst gering zu halten. Dabei reicht die Komplexität des akustischen Signals von einfachen Schwingungen bis zu Klängen mit differenzierten Spektren. Die verschiedenen Syntheseverfahren besitzen unterschiedliche Möglichkeiten zur Steuerung der Synthese.

Diese Steuerung der Erzeugung sollte einfach gestaltet sein, z. B. mit einem Satz perzeptueller Parameter für die Einstellung der Eigenschaften von Raumnachhall, damit bei der Verwendung von Klängen aus Anwendungen heraus nicht übermäßiges Verständnis für die Steuerung durch die Parameter aufgebracht werden muss (Jot 1992). Eine weitere einfache Weise, Klangeffekte zu programmieren, wurde von Beaudouin-Lafon und Gaver (1994) vorgestellt, bei der kausale akustische Zusammenhänge von Objekten und Aktionen über eine Baustruktur entwickelt werden können.

Zwei gegensätzliche Tendenzen sollten für die Gestaltung akustischer Darstellungen bei der Mensch-Computer Interaktion zusammengeführt werden. Zum einen stehen die Anforderungen von Multimediaanwendungen nach komplexen Audiosignalen derzeit im Widerspruch zur Forderung nach einfacher Integration in Anwendungen. Zum anderen sind die Methoden der Klanggeneration in der Computermusik nicht effizient und die einfache Kontrollierbarkeit der Erzeugung ist weniger ausschlaggebend, als vielmehr die Komplexität des erzeugten Klangs. Für eine akustische Darstellung sollte das Ziel angestrebt werden, gleichzeitig eine ausführliche Kontrolle über die komplexe Klangerzeugung bei der Interaktion zu realisieren, die auch ohne großen Aufwand in Anwendungen integriert werden kann.

4.3.1 Syntheseverfahren

Die grundlegenden Verfahren digitaler Synthesetechniken stellte Smith (1991) zusammen. Diese Darstellung spiegelt zum einen die historischen Ansätze aus dem Bereich der sog. elektroakustischen Musik und der Computermusik wider, zum anderen werden auch die Übergänge zu den Bereichen der Modellierung und Verfahren zur Gestaltung von Klängen deutlich. Für eine detaillierte und ausführliche Beschreibung

der Syntheseverfahren soll auf Roads (1996) verwiesen werden. Smith unterscheidet zwischen vier grundlegenden Techniken bzw. Verfahren:

- Klangaufzeichnungen,
- spektrale Modellierung,
- physikalische Modellierung und
- Algorithmen.

In den Anfängen der Computermusik wurden sowohl Verfahren der Klangaufzeichnung wie auch Algorithmen angewendet. Mit zunehmender Leistungsfähigkeit der Computerumgebungen wurden dann auch die Verfahren der Modellierung möglich, da mit ihnen aufwändige Operationen wie z. B. die Fourier-Transformation oder digitale Filterung nötig wurden. Der Grad der Abstraktion von natürlichen Klängen nimmt bei dieser Klassifikation von der einfachen Aufzeichnung bis zur Synthese mit abstrakten Algorithmen zu.

Typischer Vertreter des Verfahrens mittels Klangaufzeichnung ist der noch heute weit verbreitete Wavetable- oder Samplesynthesizer, der auch stellenweise auch Bestandteil von Multimedia-Soundkarten ist. Mit ihm ist es möglich, eine einzelne Periode einer aufgezeichneten Wellenform so zu wiederholen, dass ein kontinuierliches Signal entsteht. Darüber hinaus können auch länger aufgezeichnete Audiodaten (bis zu mehreren Sekunden) ‚geloopt‘ werden, so dass der Eindruck eines periodischen Signals entsteht. Die Verfahren mit abstrakten Algorithmen basieren auf grundlegenden Methoden der Signalerzeugung. Die Wellenformen der Generatoren sind z. B. reine Sinus-, Rechteck-, Dreieck- oder ‚Sägezahnwellenformen‘, die technisch einfach zu erzeugen sind und mit mathematischen Funktionen beschrieben werden können. Diese Generatoren können wiederum in ihrer Amplitude und Frequenz von anderen Generatoren moduliert werden, was zu deutlich komplexeren Audiosignalen führen kann. Der 1983 z. B. von der Firma Yamaha vorgestellte Synthesizer DX-7 verwendete zur Synthese die Frequenzmodulation (FM). Die sechs digitalen Sinus-Generatoren des Gerätes konnten sich in verschiedener Art gegenseitig modulieren und zusätzlich zwischen Ausgang und Eingang von Generatoren rückgekoppelt werden. Um die ‚starr‘ und ‚gefroren‘ wirkenden Klänge natürlicher erscheinen zu lassen, werden oft Klangeffekte z. B. Nachhalleffekte dem Klang nachträglich zugefügt (Smith 1991).

Die Verfahren der spektralen Modellierung bei der Synthese gehen von den Methoden der Verarbeitung von aufgezeichnetem Klangmaterial aus, werden aber durch die Möglichkeiten der Transformation des Signals in den Frequenzbereich erweitert. Dort kann es analysiert und nach anderen Gesichtspunkten als denen im Zeitbereich

manipuliert werden. Damit ist es einerseits möglich, sehr präzise die spektralen Anteile des Klangs z. B. unabhängig von seiner Grundtonhöhe zu formen. Andererseits kann damit eine zeitvariante Manipulation im Frequenzbereich vorgenommen werden. Für die spektrale Modellierung wurden auch viele Verfahren der Spektralschätzung wie die Eigenwertanalyse (principal component analysis, PCA), die lineare Prädiktion (linear prediction coding, LPC) etc. verwendet werden (Marple 1987), um z. B. stochastische Anteile von harmonischen zu trennen, unterschiedlich zu manipulieren und zu synthetisieren (Serra und Smith 1990).

Mit physikalischen Modellen wird versucht, den direkten Zusammenhang zwischen den physikalischen Eigenschaften von schwingenden Massen in Audiosignale im Zeitbereich umzusetzen. Physikalische Modellierung hat somit ein wissenschaftliches und künstlerisches Ziel (Roads 1996). Das wissenschaftliche Ziel besteht darin, in Anlehnung an das reale Instrument ein Modell zu erzeugen, das die physikalischen Verhältnisse des Instruments möglichst genau simuliert, d. h. ähnlich charakteristisch klingt. Bei der Suche nach dem Modell muss das mechanisch-akustische System möglichst vollständig erforscht werden, um ein gutes Ergebnis der Simulation zu erhalten. Das künstlerische Ziel der physikalischen Modellierung (PhM) ist es, auf der Grundlage der Modelle neue Instrumente zu erzeugen, die in der Realität weder herzustellen noch zu bespielen sind. Ein prominentes Beispiel für PhM ist die *Karpuls-Strong-Synthese*, mit der der Klang von gezupften Saiten bis zu einem Trommelschlag erzeugt werden kann (Karplus und Strong 1983).

4.3.2 Designverfahren

Neben den speziellen Verfahren, mit denen Audiosignale generiert werden, ist es auch entscheidend, nach welchen Gesichtspunkten eine Synthese entwickelt und verwendet wird. Sollen z. B. natürliche Klänge wie im Fall der PhM durch ein numerisches Verfahren simuliert werden, steht im Prinzip am Anfang des Designverfahrens die Analyse des zu simulierenden Klangs. Das Ziel des Designverfahrens ist es, mit der Synthese nicht nur einen möglichst ähnlichen Klang zu reproduzieren, sondern mit dem Modell und der Synthese auch verschiedene Attribute der Klangerzeugung zu modellieren (Gaver 1993b). So kann z. B. beim Modell der Saite einer Gitarre die Länge der Saite als Parameter eingestellt werden. Anstatt die einzelnen Partialtöne, die sich bei einer Verkürzung der Saitenlänge ergeben, explizit neu einzustellen, erlaubt die Modellierung die Kontrolle der Synthese über einen Parameter.

Wie schon im vorherigen Abschnitt angedeutet, muss das mechanisch-akustische System, das simuliert werden soll, nach physikalischen Zusammenhängen untersucht

werden, um zu verstehen, wie der Klang einer speziellen Anordnung entsteht. Dieser Vorgang kann auch für sich alleine ein Designverfahren darstellen, wenn es darum geht, eine reine Simulation zu erstellen, ohne dass das Ergebnis der Simulation für die Modellierung iterativ berücksichtigt wird. Im Verfahren, das Gaver (1993a) vorstellt, geht es aber außerdem darum, die für die Perzeption relevanten Attribute des Klangs zu identifizieren. Dazu kombiniert er das traditionelle Syntheseverfahren mit der PhM, indem er die Analyse aufteilt in einen akustischen und physikalischen Teil. Aus der akustischen Analyse erfolgt unter Berücksichtigung physikalischer Gesetze eine Modellierung, die hinsichtlich ihrer Funktion und Gültigkeit interpretiert wird (siehe Abb. 11).

Das Ziel dieses Designverfahrens ist es, durch Vereinfachungen eine Synthese zu erhalten, die einen perzeptuell überzeugenden Klang hervorbringt, der im Gegensatz zur reinen Simulation nicht mit dem Originalklang übereinstimmen muss. Dem Hörer soll das akustische Ereignis plausibel vorkommen und er soll die im Klang beinhaltete Information über das Ereignis wahrnehmen können. Gaver beruft sich dabei auf die Ansätze der ökologischen Psychologie wie in Abschnitt 3.2, auf Seite 32 beschrieben.

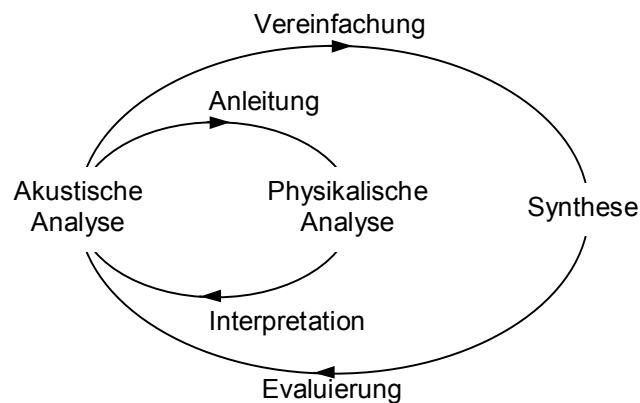


Abb. 11: Erweitertes Synthese- bzw. Designverfahren zur Gestaltung von Klängen (nach Gaver 1993a).

Andere Designverfahren lehnen sich an das von Gaver vorgestellte Verfahren an, betonen jedoch andere Aspekte. Während bei Gaver das akustische Signal eines allgemeinen Ereignisses in der Umgebung im Zentrum der Betrachtungen steht (siehe auch Abschnitt 3.2), konzentrieren sich Rauterberg et al.(1994) und Darvishi et al. (1995) auf die Interaktion zwischen Objekten mit verschiedenen Materialien. Weitere Verfahren, die stark auf perzeptuelle Effekte ausgerichtet sind, werden von sog. Foley-artists verwendet (,Klangkünstlern‘, die Effekte und Geräusche synchron zu einem Film einspielen). In heutigen Ansätzen wird versucht, dieses Verfahren zu automatisieren und

auf den Bereich der Computerspiele und -animation zu übertragen (siehe z. B. Takala und Hahn 1992, Cook 1997 und Doel 2001).

Die verschiedenen Designverfahren sollen mit den Begriffen Simulation und Imitation charakterisiert werden. Verfahren der physikalischen Modellierung haben zum Ziel, das Verhalten eines mechanisch-akustischen Systems so gut zu simulieren, dass sie einen möglichst naturgetreuen Klang generieren (physikalische Ebene). Bei der Imitation steht die plausible Vermittlung, der im Klang enthaltenen Information im Vordergrund und nicht die Richtigkeit der Modellierung. Betont wird bei der Imitation der perzeptuell relevante Anteil im akustischen Signal.

5 Neue Ansätze akustischer Repräsentationen in hybriden Umgebungen

In diesem Kapitel werden neue Ansätze akustischer Repräsentationen vorgestellt, die auf neuartige Formen der Mensch-Computer Interaktion in computerunterstützten Arbeitsumgebungen zugeschnitten sind und die die Möglichkeiten neuer Technologien berücksichtigen. Dabei liegt der Schwerpunkt der entwickelten Ansätze in der Unterstützung der Interaktion und Kooperation in sog. hybriden Umgebungen, die aus Objekten mit realen und virtuellen Eigenschaften bestehen. Mit der Forderung nach ‚hörbaren akustischen Eigenschaften‘ dieser hybriden Objekte wird der technischen Entwicklung Rechnung getragen, dass Klänge und Geräusche von Geräten zunehmend reduziert werden und sich daher Möglichkeiten eröffnen, die Mensch-Computer Interaktion mit neuartigen akustischen Repräsentationen zu unterstützen. Mit dem sog. akustischen Verhalten von hybriden Objekten wird es möglich, im anwendungs-unabhängigen Teil der Benutzungsschnittstelle dem Benutzer neuer Formen der Mensch-Computer Interaktion plausible konzeptuelle Modell anzubieten. Die neu entwickelten Ansätze werden im anschließenden Kapitel 6 bei der Realisierung von Prototypen angewendet. Es wird dabei auf die Roomware-Komponenten der hybriden i-LAND Umgebung (Streitz et al. 1999) zurückgegriffen. In dieser Umgebung wird mit neuen Formen der Mensch-Computer Interaktion für kooperative Gruppenarbeit experimentiert, bei der die Informationstechnologie in die Büroumgebung integriert ist. Beim Aufbau und der Entwicklung der i-LAND Umgebung im Bereich AMBIENTE des Fraunhofer IPSI seit 1997 wurden auch die technischen Anforderungen für akustische Repräsentationen mitberücksichtigt und damit die Grundlage für die Realisierung von Prototypen in der Umgebung geschaffen. Bei der Gestaltung der akustischen Repräsentationen spielt sowohl die Frage eine Rolle, welche Randbedingungen sich aus der Technik ergeben, als auch die Frage, wie die neuen Formen der Interaktion in solchen hybriden Umgebungen adäquat unterstützt werden können. Mit dem Begriff *Sounds@Work* soll diese Forschungsaktivität im Bereich der akustischen Repräsentationen für die Mensch-Computer Interaktion in hybriden Arbeitsumgebungen in einem prägnanten Titel abgekürzt werden.

5.1 Hybride Umgebungen

Mit dem Begriff hybride Umgebung werden solche Umgebungen bezeichnet, die *zwischen* realen Umgebungen und virtuellen Umgebungen angeordnet sind (Abb. 13). Mit dem Begriff soll darauf hingewiesen werden, dass die realen oder virtuellen Eigenschaften der Umgebungen für den Benutzer einen gleichrangigen Stellenwert

besitzen. Auf dem Kontinuum der Virtualität (Abb. 4, Seite 19) kann daher die hybride Umgebung genau in der Mitte zwischen der erweiterten Realität und der erweiterten Virtualität eingeordnet werden. Hybriden Umgebungen bestehen aus realen Objekten, die mit Informationstechnologie angereichert sind. Diese Objekte der Umgebung sollen nicht als Standardcomputer mit erweiterten Multimedia-Eigenschaften missverstanden werden, sondern die Objekte sollen zusätzlich zu ihren Funktionen im realen Raum mit weiteren Möglichkeiten der Darstellung ergänzt werden. So z. B. erhält eine Tafel durch Integration von Technik zur Erkennung der Stiftposition bei der Interaktion neue Eigenschaften, die die Verwendung der Tafel erweitern. Im Gegensatz dazu interagiert der Benutzer in einer virtuellen Umgebung mit virtuellen Objekten, wobei er z. B. durch die Verwendung von *head-mounted displays* von der tatsächlichen Umgebung getrennt ist.

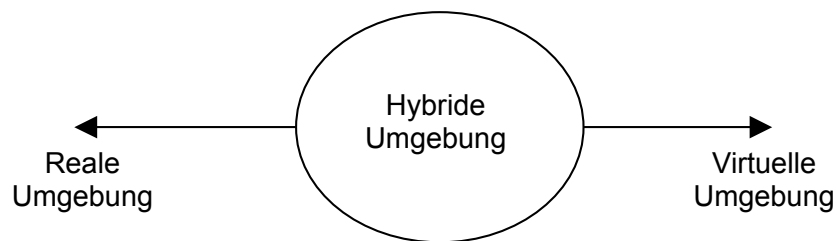


Abb. 13: Einordnung von hybriden Umgebungen zwischen realen traditionellen Arbeitsumgebungen und vollständig immersiven virtuellen Umgebungen. Bestandteil der hybriden Umgebung sind Objekte, die mit Informationstechnik erweitert sind.

Die Interaktion in einer klassischen Arbeitsplatzsituation mit Computer, Monitor, Tastatur und Maus ist standardisiert. Es existieren Interaktionsmodelle, die zur Untersuchung solcher Situationen verwendet werden können (Moran 1981, Card et al. 1983). In virtuellen Umgebungen, die die Realität simulieren (*Virtual Reality*), baut der Benutzer bei der Interaktion auf seinen Erfahrungen auf, z. B. werden seine Bewegungen analog in die virtuelle Umgebung übertragen und er erhält daraufhin eine synchronisierte multi-sensorische Darstellung (Schomaker et al. 1995). Das interne dynamische Weltmodell des Benutzers muss nicht verändert werden, um das Systemverhalten zu begreifen (Rasmussen 1986). Beim Benutzer entsteht eine visuelle und auditive Illusion, die die herkömmlichen Interaktionen ermöglicht.

In hybriden Umgebungen ist die Interaktion teilweise mit realen Objekten verknüpft, mit denen so interagiert wird, wie es in realen Umgebungen üblich ist. Durch die Erweiterung mit Computertechnik verfügen die Objekte außerdem über reale und virtuelle Eigenschaften, die neue Bedeutungen und Funktionen haben (siehe z. B. der

sog. *ambientROOM*, Ishii et al. 1998, oder das sog. *i-LAND*, Streitz et al. 1999). Diese Erweiterungen der Objekte können verschiedene Ausprägungen haben und reichen von Computersystemen mit optischen interaktiven Displays bis zu einfachen Ausstattungen zur Identifikation von Objekten mit entsprechender Sensortechnologie. Die Interaktion mit diesen Objekten erhält dadurch einen hybriden Charakter, der z. B. durch die Weiterentwicklung von grafischen Benutzungsoberflächen hin zu sog. *greifbaren* Benutzungsschnittstellen (*tangible user interface*, TUI) unterstrichen wird (Ullmer 2000). Die Interaktion kann ihren hybriden Charakter dann erhalten, wenn der konkrete physikalische Gegenstand der Interaktion zusätzlich eine Bedeutung im virtuellen Raum besitzt, aber auch, wenn das Objekt, das den Gegenstand der Interaktion bereitstellt, sowohl reale als auch virtuelle Eigenschaften besitzt. Mark Weiser nannte die Vorgehensweise, den Computer in die ‚physikalische Welt‘ zu integrieren, „*embodied virtuality*“ (verkörperte Virtualität), um den Unterschied zu ‚Virtual Reality‘ bzw. virtuellen Umgebungen hervorzuheben (Weiser 1991; siehe auch Abschnitt 2.3.2, Seite 20).

Ziel der hybriden Umgebung ist, durch die Integration realer und virtueller Anteile und Eigenschaften eine kohärente mentale Repräsentation beim Benutzer zu erzeugen. Es bleibt somit dem Benutzer überlassen, welchen Anteil er in Abhängigkeit der Situation benutzt. Akustische Repräsentationen werden verwendet, um dieses Ziel zu erreichen, d. h. ein kohärentes mentales Modell beim Benutzer zu erzeugen und darüber hinaus weitere Handlungsmöglichkeiten zu eröffnen. Auf der physikalischen Ebene der Interaktion ist es möglich, dass akustische Signale mit natürlichem Ursprung sich mit künstlich erzeugten Signalen vermischen und dadurch neue Hörerlebnisse entstehen. Während hingegen die Verbindung zwischen einem realen Objekt und seinen akustischen Eigenschaften in der Regel starr ist, kann der Anteil virtueller akustischer Eigenschaften eines Objekts frei gewählt und mit der Zeit und in Abhängigkeit der Situation variiert werden. Damit eröffnen sich weitere Gestaltungsmöglichkeiten von personalisierten und situationsabhängigen akustischen Darstellungen bei der Interaktion in hybriden Umgebungen.

5.2 Forderung nach hörbaren Objekteigenschaften

Die Entwicklungen im Bereich der Computertechnologie und die Möglichkeiten für die Verarbeitung und Darstellung von akustischen Signalen in Computerumgebungen wurden in Kapitel 4 dargestellt. Zu diesen Eigenschaften kommen aber noch diejenigen hinzu, die das akustische Verhalten des Computers bzw. der Objekte an sich ausmachen. Die typischen Geräusche in einer Computerumgebung wie die der Festplatte, des Lüfters etc. werden unbewusst vom Benutzer ausgeblendet und nur dann

Gegenstand der Aufmerksamkeit, wenn sie sich signifikant verändern, z. B. unvermittelt aufhören. Es stellt sich somit die Frage, wie mit solchen fehlenden Geräuschen zukünftiger Computer bzw. Geräte bei der Interaktion umgegangen werden soll. Eine Antwort auf diese Frage ist die Forderung nach *hörbaren Objekteigenschaften*: Damit soll erreicht werden, dass die akustischen Eigenschaften von hybriden Objekten wieder hörbar gemacht werden, um die Ausbildung mentaler Modelle dieser Objekte beim Benutzer für die Interaktion zu unterstützen.

Die Geräuschunterdrückung bzw. -vermeidung im Sinne einer Lärmverminderung in der unmittelbaren Umwelt des Menschen ist sicherlich eine zu begrüßende Entwicklung in der sonst steigenden Umweltbelastung durch unerwünschte Geräusche (Wrightson 2000). Für den Benutzer kann aber die ‚Stille‘ der Geräte und Objekte auch zu Irritationen und Missverständnissen bei der Interaktion führen, da die unbewusst erlernten akustischen Signale das allgemeine mentale Modell so geprägt haben, dass z. B. ihr Ausbleiben als Fehlfunktion interpretiert werden kann. Zum einen wird diese Aussage durch die Erfahrungen gestützt, die der Autor bei der Entwicklung und der Erprobung von geräuschlosen Computersystemen gemacht hat. Zum anderen fordert z. B. Engelen von Philips Design (1998) für das Geräuschverhalten von Haushaltsgeräten, dass die Geräusche derartiger Geräte abgewandelt und angenehmer gemacht sollten. Das einzige Ziel in der Geräuschverminderung zu suchen, ist ihrer Meinung nach für die Gestaltung von Geräten nicht ausreichend: „Silence is not always the best solution” (Engelen 1998). Beispiele akustischer Signale, die technisch betrachtet nicht nötig sind und mittlerweile vermieden werden können, sind die Übersprechgeräusche beim ‚Wählen‘ bzw. Tippen einer Telefonnummer und die Schaltgeräusche des Kfz-Blinkerrelais. In beiden Fällen dieser akustischen Rückkopplung ist der originäre Grund des Geräuschs durch Entwicklungen im technischen Bereich nicht mehr gegeben. Vielmehr wird ein Geräusch künstlich hinzugefügt, um erlernten Mustern akustischer Geräteigenschaften zu genügen. Als prominentes und aktuelles Beispiel ist auch der Verschlussklang von digitalen Fotokameras zu nennen, der bei bestimmten Modellen frei wählbar ist.

Mit der Forderung nach hörbaren Objekteigenschaften ist nicht nur gemeint, dass das Gerät bei der Interaktion akustische Signale produzieren soll, sondern auch, dass die akustischen Signale des Geräts in Abstimmung mit dem Kontext der Verwendung gestaltet werden. Für die Benutzungsschnittstelle spielen die akustische Darstellung von Objekten und deren Zustände auf der Systemoberfläche eine ähnliche Rolle wie optische Darstellungen. Gerade in hybriden Umgebungen, in denen die Systemoberfläche nicht standardisiert ist, sind dann die besonderen Eigenschaften von akustischen gegenüber optischen Darstellungen ein zusätzliches wichtiges

Gestaltungsmittel. Geht man von geräuschlosen hybriden Objekten aus, können bei der Gestaltung der hörbaren Objekteigenschaften verschiedene Wege beschritten werden:

- Die Geräusche, die nicht mehr physisch produziert werden, können wieder an die ‚Systemoberfläche‘ zurück projiziert werden und dort in Abhängigkeit von z. B. Präferenzen des Benutzers und der Bedeutung des Objekts dargestellt werden. Damit können vorhandene individuelle mentale Modelle des Benutzers bedient werden.
- Neue zusätzliche Eigenschaften der hybriden Objekte können an der Systemoberfläche hörbar gemacht werden. Für sie bestehen bei Benutzern keine Erwartungen und die Bedeutung der akustischen Signale müssen vermittelt bzw. erlernt werden.

Der Computer, der z. B. durch Miniaturisierung und Integration physikalisch ‚unsichtbar‘ geworden ist (*disappearing computer*), muss in seinen essentiellen Funktionen und Bedeutungen jedoch für die Interaktion über eine Darstellung verfügbar bleiben. Die Forderung nach hörbaren Objekteigenschaften kann daher auch als komplementär zur Forderung nach sichtbaren Eigenschaften von Darstellungen verstanden werden, damit überhaupt Handlungsmöglichkeiten entdeckt werden können (Gaver 1991b).

5.3 Akustisches Verhalten von hybriden Objekten

Das *akustische Verhalten* von hybriden Objekten bei der Mensch-Computer Interaktion kann in zwei Teile zerlegt werden, wobei unter Verhalten hier Änderungen im Zustand des Objektes bzw. Systems verstanden werden sollen. Der eine Anteil ist abhängig vom Sachverhalt der Interaktion bzw. der Anwendung, während der andere Anteil unabhängig von der Anwendung ist und dem Prozess der Interaktion zugeordnet ist (Interaktionsproblem, Streit 1985). Häufig wird bei der Mensch-Computer Interaktion ein anwendungsabhängiges akustisches Verhalten realisiert. Im Fall der Interaktion in hybriden Umgebungen soll mit anwendungsunabhängigen Rückkopplungen die Gerät- bzw. Objekteigenschaft unterstützt werden, um die Bildung der kohärenten Repräsentation zu unterstützen.

Ein Beispiel für ein anwendungsunabhängiges akustisches Verhalten in traditionellen Computerumgebungen ist z. B. das Klicken beim Drücken und Loslassen einer Taste auf der Computermouse. Der ‚Klick‘ ist die Reaktion des Objekts Maus auf die Einwirkung des Benutzers auf das Objekt. Unbewusst sicherlich oft bei der Benutzung ausgeblendet, vermittelt der Klick unzweifelhaft die Bestätigung der intendierten

Interaktion, unabhängig von der speziellen Situation in der aktuellen Anwendung. In manchen Programmen wie z. B. LexiROM 4.0 der Firma Microsoft wird die Auswahl aus dem Menü zusätzlich durch kurze akustische Signale über die Lautsprecherausgabe am Computer dargestellt, die dann nahezu gleichzeitig mit den Geräuschen der gedrückten Maustaste hörbar werden. Es wird hier schon deutlich, dass mit dem Computer die Möglichkeit geschaffen wird, ein gewohntes akustisches Verhalten von Objekten bei der Interaktion zu imitieren. Das Verhalten dient zur Bestätigung der Interaktion und vermittelt durch Invarianz gleichzeitig die Existenz des Geräts. Der Ursprung des akustischen Signals kann aber auch im Gerät bzw. Objekt selbst liegen und damit die Funktion der Benachrichtigung besitzen, z. B. die Änderung eines Systemzustands darstellen. Jenseits der speziellen anwendungsabhängigen akustischen Signale soll mit dem anwendungsunabhängigen akustischen Verhalten von Objekten in hybriden Umgebungen die Voraussetzung geschaffen werden, die Interaktion und Kooperation zu unterstützen.

5.4 Technische Aspekte hybrider Umgebungen

Zweifellos ist die Verfügbarkeit von steigender Prozessorleistung bei immer kleiner werdenden Abmaßen der Computerkomponenten mit ausschlaggebend für die Entwicklung von hybriden Umgebungen. Das ‚Gesetz‘ nach Gordon Moore aus dem Jahre 1965, nach welchem sich ca. alle zwölf Monate die Speicherkapazität von Siliziumchips verdoppelt, kann zugleich als Motor und Indikator der Entwicklung angesehen werden. 1973 korrigierte Gordon das Gesetz und erhöhte die Zeitdauer der Verdopplung auf ca. 18 Monate. Das Gesetz hat noch heute Gültigkeit und es erlaubt, Voraussagen über zukünftige technische Entwicklungen zu treffen. So lässt sich damit der Zeitpunkt bestimmen, an dem spezielle Hardwareentwicklungen für aufwändige Berechnungen durch reine Softwarelösungen auf Standardcomputern in ihrer Leistungsfähigkeit überboten werden. Mit den kleiner werdenden Abmaßen der Informationstechnik und die zunehmende Verbreitung drahtloser Netzwerkverbindung wird die Unterstützung der Mobilität des Benutzers möglich. Vor dem Hintergrund der akustischen Darstellung ist es wichtig, dass in der Regel auch die Ein- und Ausgabe von Audiosignalen zur Ausstattung von sog. PDAs gehört. Damit existiert für viele Anwendungen, die auf der Mobilität des Benutzers aufbauen, eine Plattform, mit deren Hilfe neuere Ansätze der Mensch-Computer Interaktion mit akustischen Darstellungen einfach zu realisieren und aufzubauen sind (Mynatt et al. 1998, Sawhney und Schmandt 1999). Gleichzeitig stellt die Möglichkeit der drahtlosen Netzwerkanbindung dieser Geräte der PostPC-Ära eine weitere Eigenschaft dar, mit der eine generelle Geräteklasse entsteht, die viele Anforderungen hybrider Umgebungen erfüllt. Mit einer solchen

Geräteklasse wird auch die Konvergenz zwischen den Produkten der Informationstechnologie und denen der Telekommunikation immer deutlicher: z. B. ein PDA, mit dem man auch telefonieren kann, versus einem Mobiltelefon mit persönlichem Kalender und weiteren Anwendungen.

Weitere Technologien, die besondere Bedeutung für hybride Umgebungen haben, sind solche, die die Interaktion betreffen und ermöglichen. Im Vordergrund hierbei stehen die Entwicklungen von berührungsempfindlichen Oberflächen, die gleichzeitig die Projektionsebenen der optischen Darstellung sind. Diese sog. Touchscreens erlauben in der physikalischen Ebene des Interaktionsmodells eine direkte manuelle Kontrolle der Elemente auf der Darstellungsoberfläche des Computers. Zu der minimierten semantischen Distanz zwischen der Repräsentation auf der Systemoberfläche und dem konzeptuellen Modell, das durch das Prinzip der direkten Manipulation entsteht (Shneidermann 1992), wird durch die Technologie der Touchscreens zusätzlich noch die physikalische Distanz zwischen dem Ort der Eingabe und dem Ort der Exekution bzw. Ausgabe minimiert. Dadurch wird ähnlich wie in virtuellen Umgebungen eine Illusion durch eine analoge Repräsentation bei der Interaktion erwirkt. Die technische Variante dieser Art Interaktion, bei der ein spezieller Stift benötigt wird, besitzt den Vorteil, dass weitere Parameter über die Interaktion wie Druckstärke und Anstellwinkel des Stiftes usw. zur Verfügung stehen. Das Auflegen des Handballens auf die Interaktionsfläche zur Unterstützung der Handführung bei der Interaktion führt bei dieser Variante zu keiner Fehldetektion bzw. -interpretation, da nur der Ort der Stiftspitze für die Interaktion entscheidend ist. Darüber hinaus muss der Stift für die Detektion die Oberfläche nicht berühren und mehrere Stifte können gleichzeitig bei der Interaktion unterschieden werden. Diese Technologie aus dem Bereich der sog. Graphikdigitizer ist auch die Grundlage der neuen Klasse sog. Tablet-PC's, tragbarer Geräte mit Stifteingabe.

6 Realisierungen von akustischen Repräsentationen für die i-LAND Umgebung

Im Rahmen dieser Arbeit wurde die i-LAND Umgebung zur kooperativen Gruppenarbeit (Streitz et al. 2001) als Beispiel einer hybriden Umgebung benutzt, um zu untersuchen, welche Möglichkeiten für akustische Repräsentationen in solchen Umgebungen bestehen und um Prototypen exemplarisch aufzubauen. Im Zentrum der i-LAND Umgebung stehen sog. Roomware-Komponenten, mit denen der Weg beschritten wird, Informationstechnik in existierende Elemente eines Raums wie Tisch, Stuhl, Wand etc. zu integrieren (siehe Abb. 15). Dieser Gestaltungsansatz verleiht den Roomware-Komponenten die Qualität von hybriden Objekten: Weder handelt es sich bei den Roomware-Komponenten um traditionelle Computerarbeitsplätze mit Monitor, Tastatur und Maus noch sind sie Bestandteil einer rein virtuellen Umgebung. Über zweidimensionale interaktive optische Darstellungen an den Roomware-Komponenten, durch die sich den Benutzern ein ‚virtueller Arbeitsraum‘ eröffnet, kann intuitiv direkt an elektronischen Dokumenten gearbeitet werden. Sämtliche Bildpunkte in der i-LAND Umgebung sind interaktiv und können entweder direkt mit dem Finger oder mit einem speziellen Stift manipuliert werden.



Abb. 15: Die i-LAND Umgebung mit den Roomware-Komponenten ConnecTable, CommChair, InteracTable (v.l.n.r.) und DynaWall (im Hintergrund).

Die Roomware-Komponenten bilden in einem physikalischen Raum angeordnet eine Arbeitsumgebung, die von einer Gruppe für Besprechungen, Vorführungen und

Präsentationen genutzt werden kann. Der zusätzliche virtuelle Arbeitsraum, der über die interaktiven optischen Darstellungen auf den Komponenten für die Benutzer zugänglich ist, erweitert die reale Arbeitsumgebung um die Möglichkeit zur Kooperation in einem virtuellen Arbeitsbereich. Die Szenarien der Gruppenarbeit in der i-LAND Umgebung sind auf synchrone Kooperation am gleichen Ort ausgerichtet (siehe dazu Steinmetz 2000) und zielen auf die Unterstützung kreativer und dynamischer Gruppenarbeit. Die direkte, spontane und informelle Kommunikation innerhalb der Gruppe mittels Sprache unterliegt in dieser hybriden Umgebung den gleichen Bedingungen wie in normalen Räumen. Dadurch geht der Face-to-Face Kontakt (eine „reiche“ und interaktive Kommunikationsform nach Kraut et al. 1990) in der i-LAND Umgebung nicht verloren und die Umgebung eignet sich daher zur Bearbeitung von Aufgaben mit hoher Mehrdeutigkeit, die sich nicht durch viel Information lösen lassen. Diese Aufgaben unterliegen vielmehr der Interpretationsfähigkeit der Mitarbeiter, die zu einem gemeinsamen Verständnis eines Sachverhalts kommen müssen (Schwabe 2001). Für die Unterstützung der Aufgaben mit hoher Mehrdeutigkeit wird die Form der direkten, Face-to-Face Kommunikation gefordert.

Die Softwareumgebung BEACH (*Basic Environment for Active Collaboration with Hypermedia*, Tandler 2001) ermöglicht die verteilte Kooperation zwischen Benutzern von verschiedenen Roomware-Komponenten aus, so dass z. B. an der DynaWall dargestellte Objekte auch vom CommChair aus bearbeitet werden können. In *private workspaces* ist es den Benutzern der Roomware-Komponenten wiederum möglich, auch nicht-öffentliche, private Notizen zu machen. Die Software gibt keine besonderen Einschränkungen oder Regeln für die Kooperation in der i-LAND Umgebung vor, um die kreativen Prozesse bei der Gruppenarbeit nicht zu behindern. Kreativitätstechniken wie Brainstorming oder Metaplan werden unterstützt und generische Funktionalitäten wie z. B. die Visualisierung von Wissensstrukturen können verwendet werden (Streitz et al. 2001). Technisch möglich sind auch verteilte Szenarien zur Kooperation zwischen einer Gruppe und weiteren Personen an einem entfernten Ort oder die Kooperation von zwei Gruppen an verschiedenen Orten.

6.1 Einordnung akustischer Repräsentationen

Bevor in den nachfolgenden Abschnitten die Beispiele akustischer Repräsentationen beschrieben werden, sollen sie eingeordnet werden. Die drei Kategorien von Funktionen akustischer Darstellungen, wie sie für die Interaktion mit Standardcomputern identifiziert wurden: die Rückkopplung, die Benachrichtigung von Systemzuständen und die Vermittlung des Zustands der Interaktion und der Aktivität anderer Benutzer (Beaudouin-Lafon und Gaver 1994), können dafür als Ausgangspunkt genommen

werden. Die realisierten Prototypen in dieser Arbeit lassen sich folgende Bereiche aufteilen:

- die Anreicherung der Interaktion durch akustisches Feedback,
- die akustische Repräsentation von Systemzuständen und
- die hybride Präsentation von akustischen Ereignissen.

Der *Grad der Erweiterung* durch eine akustische Repräsentation soll als ‚schwach‘ bezeichnet werden, wenn dem hybriden Objekt lediglich Eigenschaften hinzugefügt werden, die dem Benutzer schon bekannt sind und die er aus dem Umgang mit realen Objekten erwartet.

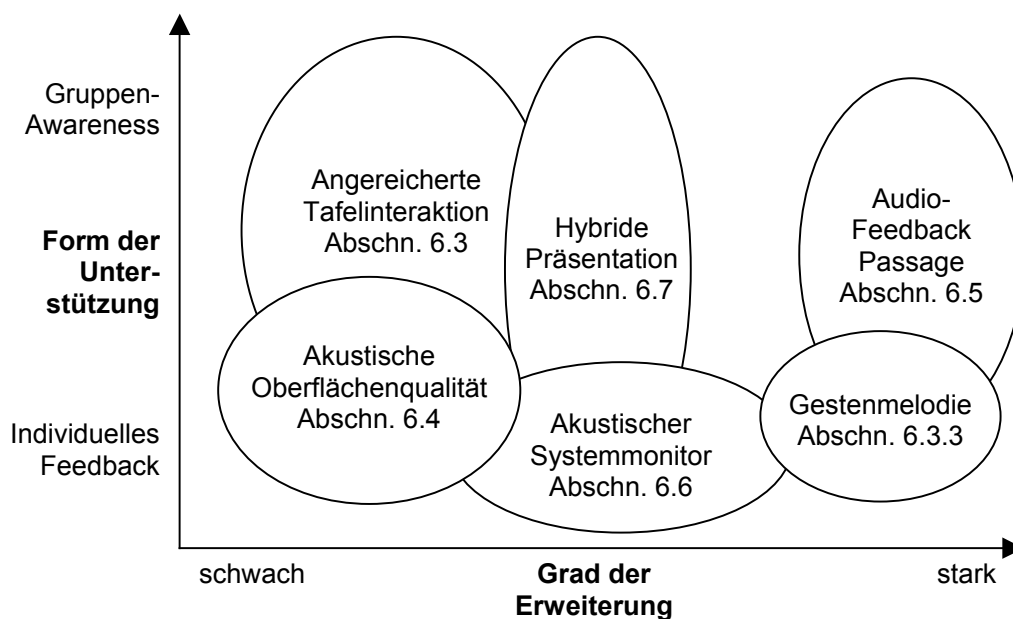


Abb. 16: Klassifikation der verschiedenen Beispiele akustisch angereicherter Interaktion mit der hybriden Testumgebung i-LAND.

Der Grad der Erweiterung wird als ‚stark‘ bezeichnet, wenn dem Benutzer bei der Interaktion neue Eigenschaften hybrider Objekte angeboten werden. Ähnlich wie bei einem zusätzlichen, neuen Schalter an einem Gerät muss der Benutzer dann die Bedeutung und Funktion erlernen. In Abb. 16 sind die verschiedenen Beispielanwendungen, die in dieser Arbeit realisiert wurden und nun in diesem Kapitel beschrieben werden, dargestellt. Sie werden klassifiziert nach ihrem Grad der Erweiterung und der Unterstützung für Individuen bzw. Gruppen. Es wird dabei zwischen einer individuellen akustischen Darstellung für eine Person und einer akustischen Darstellung für Gruppensituationen (Gruppen-Awareness) unterschieden. Eine akustische Darstellung ermöglicht je nach Betrachtungsperspektive beide Formen

der Unterstützung. Grund dafür ist die schon in Kapitel 3 angedeutete Besonderheit akustischer Signale, die darin besteht, dass akustische Ereignisse den Raum erfüllen und dadurch als eine öffentliche Darstellung bezeichnet werden kann: „The problem here is that audio is an inherently public medium“ (Hindus et al. 1995, Seite 145). Mit der Form der Unterstützung sollen in diesem Zusammenhang nur Eigenschaften der akustischen Repräsentation berücksichtigt werden, die speziell durch die Verwendung hybrider Objekte entstehen.

6.2 Gestaltungsregeln für akustische Darstellungen

Verschiedene Regeln für die Gestaltung der akustischen Darstellungen wurden aufgestellt, um bei der Erweiterungen der Interaktion in der hybriden i-LAND Umgebung einen einheitlichen Charakter zu erhalten. Es soll damit die Grundlage geschaffen werden, den Ansatz der Integration realer und virtueller Eigenschaften auch im akustischem Medium zu ermöglichen. Die Regeln sind:

- A) die Verwendung von ‚unsichtbaren‘ Darstellungsmittel in der Umgebung,
- B) die Erzeugung der akustischen Signale an dem Ort, an dem sie für die Interaktion entscheidend sind,
- C) die Erzeugung der akustischen Signale unter Verwendung mehrerer Informationen über die Interaktion und über die hybride Umgebung und
- D) die sofortige und direkte Rückkopplung der Interaktion mit hybriden Objekten über die akustische Darstellung.

Die Regel A zielt darauf ab, zu verhindern, dass die Aufmerksamkeit des Benutzers bei der Interaktion auf sichtbare Lautsprecher und Kabel in der hybriden Umgebung gelenkt wird, die keinen hybriden Charakter haben. Durch die Präsenz von Lautsprechern kann der intendierte auditive Eindruck beeinträchtigt werden. Eine einfache Anordnung mit Lautsprechern auf dem Tisch, wie sie bei Multimedia-Computern üblich ist (z. B. von Gaver 1991a verwendet) sollte daher vermieden werden. Regel B sollte erfüllt sein, damit der Benutzer bei der Interaktion nicht durch eine räumliche Inkonsistenz zwischen dem Ort der Interaktion und dem Ort der akustischen Repräsentation irritiert wird. Der Eindruck, dass ein Objekt eine zusätzliche Eigenschaft bzw. akustische Darstellung besitzt, wird zusätzlich verfestigt, wenn der Ort der Interaktion dem der räumlichen Anordnung der akustischen Darstellung entspricht. Um die Probleme der Wahrnehmung einer virtuellen Schallquelle durch die Projektion mehrerer Lautsprecher zu vermeiden, ist es weiterhin sinnvoll, nahe am Ort der Interaktion auch einen Schallwandler zu platzieren. Die Regel C soll den ansonsten üblichen Ansätzen in Computerumgebungen entgegentreten und helfen, das starre Abspielen von

aufgezeichneten akustischen Signalen zu vermeiden. Je mehr Informationen über die Interaktion mit einem hybriden Objekt zur Verfügung stehen, um so plausibler wird der Höreindruck durch die akustische Darstellung. Damit Probleme der zeitlichen Integration eines oder mehrerer Ereignisse durch verschiedene Modalitäten vermieden werden, ist die Forderung D nach sofortiger Rückkopplung schließlich äußerst sinnvoll. Soll eine zusätzliche virtuelle Eigenschaft eines Objekts beim Benutzer zu einer kohärenten mentalen Repräsentation führen, muss bei der Vermittlung dieser Eigenschaft Aspekte der Auflösungsschwelle und der Ordnungsschwelle berücksichtigt werden (siehe Abschnitt 3.1.1, Seite 26).

6.3 Unterstützung von Interaktion an der DynaWall

Die Interaktion an einer traditionellen Wandtafel ist die Grundlage für die Interaktion an der DynaWall. Der Benutzer kann mit einem Stift oder aber auch mit dem Finger, Elemente auf der DynaWall erzeugen oder mit ihnen interagieren. Bei dieser Form der direkten Manipulation wird neben der semantischen Distanz zwischen der Repräsentation auf der Tafeloberfläche und dem der Darstellung zugrunde liegenden konzeptuellen Modell zusätzlich die physikalische Distanz bei der Interaktion verkleinert (siehe Abschnitt 5.4, Seite 56). Die Repräsentationen können an der DynaWall ohne zusätzliche Hilfs- und Zeigemittel, wie z. B. einer Computermouse, direkt mit dem Finger ‚berührt‘ werden und stellen damit eine Oberfläche dar, die über reale und virtuelle Eigenschaften durch die Computerprojektion verfügt.

In Erweiterung zur traditionellen Tafel, auf der lediglich mit einem Stift geschrieben oder gezeichnet werden kann, ist es auf der DynaWall möglich, besondere Bereiche (*workspaces*) zu schaffen, in denen geschrieben oder gezeichnet werden kann und die überall auf der Tafeloberfläche platziert werden können. Mit dieser und weiteren Eigenschaften übersteigt die DynaWall deutlich die Merkmale traditioneller Tafeln. Die Verwendung der Tafelmetapher ist darüber hinaus entscheidend für den Charakter der Situationen, in denen an der DynaWall gearbeitet wird. Durch die Größe der Darstellungen an der DynaWall ist es möglich, sie für Gruppensituationen wie Vorträge und Präsentationen oder Kleingruppenarbeit zu verwenden. Obwohl die Darstellung an der DynaWall auch nur eine Ausdehnung in zwei Dimensionen besitzt, ist die Art der Nutzung nicht zu vergleichen mit der eines Standardcomputermonitors für einen Einzelarbeitsplatz. Die räumliche Ausdehnung der DynaWall bewirkt, dass aus verschiedenen Perspektiven auch verschiedene Ansichten der Oberfläche der elektronischen Tafel für den Benutzer entstehen, so dass die Oberfläche teilweise sogar nicht vollständig sichtbar ist (siehe auch Abb. 17).



Abb. 17: Kooperative Arbeitssituation an der DynaWall: Die Person links im Bild verschiebt mit dem Finger ein Element auf der Oberfläche. Diese Person hat eine andere Perspektive auf die Darstellung an der Oberfläche als die Person rechts im Bild.

Die Unterstützung der Interaktion an der DynaWall durch akustisches Feedback zielt auf zwei Bereiche der Interaktion ab:

- die Unterstützung und Bereicherung individueller Interaktion mit der DynaWall und
- die Unterstützung einer Gruppe bei der Kooperation an der DynaWall.

Die DynaWall wird unter Berücksichtigung der Gestaltungsregeln A und B um eine zusätzliche akustische Darstellung erweitert. Damit können für die optischen Objekte auf der Oberfläche an der DynaWall zusätzlich akustische Eigenschaften erzeugt werden, die die Benutzerinteraktion begleiten und die Erwartungen des Benutzers bei der Interaktion mit den Objekten erfüllen.

6.3.1 Akustische Darstellung für die Interaktion an der DynaWall

Die akustische Darstellung von Objekten an der DynaWall ist so konzipiert, dass diese Darstellung die Eigenschaften der optischen Darstellung ergänzt und dadurch die Voraussetzung für eine multimodale Interaktion an der elektronischen Tafel geschaffen wird. Die Darstellungsarten unterscheiden sich jedoch deutlich in ihrem zeitlichen Verhalten, wobei die optisch wahrnehmbaren Objekte die Zeit ‚überdauern‘, wohingegen akustische Ereignisse eher flüchtig sind und mit der Zeit vergehen (Gaver 1997). Kramer (1994) nennt außerdem zusätzliche Vorteile akustischer Darstellungen in Verbindung mit optischen Darstellungen, wie z. B. die höhere zeitliche Auflösung und die Erhöhung der wahrgenommenen Darstellungsqualität. Die akustische Darstellung an der DynaWall zielt weniger auf die Repräsentation des Inhalts eines bearbeiteten Dokuments ab, sondern vielmehr auf die Repräsentation der Interaktion mit dem Dokument (Interaktionsproblem, Streit 1985), wenn z. B. durch Interaktion Änderungen an den Objekten vorgenommen werden. Die Bereiche, bei denen die Interaktion durch eine akustische Darstellung angereichert wird, sind:

- die Bearbeitung eines Objektes mit einem Stift (Stift-Annotation),
- die kontrollierte Positionsänderung eines Objektes und
- die automatischen Bewegungsabläufe von Objekten.

Mit Stift-Annotationen sind Markierungen gemeint, die ähnlich wie handschriftliche Notizen auf Dokumenten benutzt werden und im Wesentlichen die Funktionen und Möglichkeiten eines einfachen Schreibstifts aufweisen. Der Fall der kontrollierten Positionsänderung eines Objektes entspricht in etwa der Interaktion mit der Computermouse, bei der z. B. ein Datei-Ikon oder ein ‚Fenster‘ auf der Desktop-Oberfläche verschoben wird. Entsprechend wird das Objekt an der DynaWall ‚berührt‘ und auf der Oberfläche bewegt. Automatische Bewegungsabläufe werden an der DynaWall immer dann ausgelöst, wenn ein Objekt der Darstellung durch die Interaktion des Benutzers eine deutliche Beschleunigung erfährt und dann plötzlich ‚losgelassen‘ wird. Danach vollzieht das Objekt eine Bewegung in die Richtung und mit der Geschwindigkeit, die es zuletzt bei der direkten Kontrolle durch den Benutzer besaß. Dieses Verhalten ist angelehnt an das von realen Objekten, die z. B. geworfen werden, wobei an der DynaWall die Bewegung auf zwei Dimensionen begrenzt bleibt. Die akustische Unterstützung des ‚Werfens‘ von Objekten entwickelt sich aus dem Geräusch der kontrollierten Positionsänderung und erweitert dadurch die invarianten Eigenschaften der Oberfläche.

Die kontrollierte Interaktion mit Objekten an der DynaWall entspricht denen von adhäsiven Objekten auf Oberflächen. Eine bestimmte Ausrichtung der Oberfläche, z. B. horizontal oder vertikal, spielt dabei keine Rolle. Bei automatischen Bewegungen von Objekten verhält sich die vertikale Darstellung jedoch so wie eine horizontale z. B. auf einem Tisch. Die automatische Bewegung der Objekte hängt nicht von einer ‚Gravitationskraft‘ ab, die auf die Objekte wirkt, sondern lediglich von der Benutzerinteraktion. Die zusätzliche Berücksichtigung einer solchen virtuellen Kraft würde zwar realem Verhalten näher kommen, ist aber für die Benutzerinteraktion nicht hilfreich. Das konzeptuelle Modell sog. ‚masseloser‘ Objekte erfüllt daher nicht die Erwartung des Benutzers. Diese Unvollständigkeit des konzeptuellen Modells kann aber durch das zusätzliche Angebot der akustischen Anreicherung kompensiert werden.

Ein weiterer Bereich, in dem spezielle Formen der Interaktion an der DynaWall durch akustische Darstellungen unterstützt werden, wird in Abschnitt 6.3.3 beschrieben. Mit speziellen Gesten können bei der Interaktion Elemente auf der Oberfläche erzeugt, gelöscht etc. werden. Im Sinne der Definition multimodaler Interaktion (siehe Abschnitt 2.2.2) wird hier aus der Information der Interaktion eine Bedeutung gewonnen, die in diesem Fall als Befehle an die Objekte interpretiert werden.

6.3.2 Virtuelle Reibung

Die akustische Unterstützung bei der Interaktion mit Objekten auf der Oberfläche der DynaWall ist i. A. mit dem oben beschriebenen optischen Verhalten bei der Interaktion verknüpft. Das Konzept, das der akustischen Darstellung zugrunde liegt, ist das der *virtuellen Reibung* zwischen dem Objekt und der Oberfläche. Die Reibung zwischen einem Objekt und der darunter liegenden Oberfläche wird dem Benutzer bei der Interaktion über die akustische Darstellung vermittelt. Diese Reibung wird auch offensichtlich, wenn Objekte automatisch Bewegungsabläufe ausführen, da dann ihre Geschwindigkeit gemäß einer Oberflächenreibung kontinuierlich vermindert d. h. negativ beschleunigt wird. Das Konzept der virtuellen Reibung kann im Zusammenhang mit der optischen Darstellung auf der Systemoberfläche dem Benutzer gut vermittelt werden und unterstreicht die Beziehung zwischen Objekten und der Oberfläche.

Bei der Kontrolle der Position von Objekten der optischen Darstellung fließt an der DynaWall hauptsächlich die Geschwindigkeit der Interaktion in die akustische Darstellung ein. Die Kraftgrößen spielen für die akustische Darstellung an der DynaWall, wie oben schon erwähnt, eine geringere Rolle. Der Ausschnitt einer realen Aufnahme eines Gegenstands, der auf einer Tischoberfläche verschoben wird, bildet die Grundlage für diese Darstellung der Interaktion (kontrollierte Positionsänderung).

Werden darüber hinaus automatische Bewegungsabläufe initiiert, unterstützt ein zusätzliches Resonanzfilter den Eindruck der abgebremsten Bewegung. Das Filter wird in seiner Eckfrequenz und Resonanzgüte durch die langsamer werdende Geschwindigkeit des bewegten Objektes kontrolliert.

Bei der Interaktion in Form der Stift-Annotation wird mit ‚Kreide‘ bzw. mit einem realen Stift (jedoch mit virtueller Tinte) auf Objekten an der DynaWall interagiert. Die Grundlage des Klangs bei dieser Interaktion liefert ein Ausschnitt einer realen Aufnahme eines Kreidestrichs auf einer Tafeloberfläche. Beim Beginn der Interaktion wird zusätzlich das Aufsetzen der ‚Kreide‘ auf der Tafeloberfläche durch ein kurzes, abklingendes Geräusch imitiert. Bei zeitlich anhaltender Interaktion wie z. B. beim Zeichnen langer Linien wird die akustische Darstellung langsam nach einigen Sekunden ausgeblendet, um eine zu starke Präsenz anhaltender Geräusche zu vermeiden. Gaver berichtet von einer Interviewstudie, in der Personen das Geräusch sehr gut wiedererkennen können, das entsteht, wenn mit Kreide auf einer Tafel geschrieben wird (Gaver 1993a, Seite 19): „Participants did find that some sounds were extremely difficult to identify (e.g., the sound of a file drawer being opened and closed), but they were almost always correct about some others (e.g., the sound of writing with chalk on a chalkboard)“.

6.3.3 Akustische Unterstützung von Gesteninteraktion

Neben den Interaktionen, die direkte analoge Bedeutung haben, wenn z. B. ein Objekt verschoben wird oder eine Linie gezeichnet wird, können die Interaktionen auch eine weitere spezielle Bedeutung besitzen. Diese sog. Gesten sind zeitliche und räumliche Interaktionsfiguren (Trajektorien), die eine Intention des Benutzers ausdrücken. In Verbindung mit Objekten, auf denen die Geste erfolgt, wird eine Operation ausgeführt, z. B. das Kopieren eines Objektes, wie es in der Abfolge in Abb. 18 dargestellt ist. Mit dieser Funktionalität, die Teil der BEACH Software ist (Tandler 2000), wird die stiftbasierte Benutzungsschnittstelle *mode-less*, d. h. der Benutzer muss nicht explizit von einem Eingabezustand in einen Kontrollzustand des Systems wechseln, um z. B. Operationen wie Kopieren, Löschen etc. auf Operanden der Darstellung auszulösen. Diese Unterscheidung zwischen einer Eingabe oder der Steuerung der Standardanwendung muss vom Benutzer üblicherweise explizit vorgenommen werden. In der BEACH Software wird dieser Wechsel implizit durch die Benutzerinteraktion vorgenommen. Während einer allgemeinen Eingabe wird inkrementell die Interaktion auf eine potenzielle Geste hin überprüft und kann nach ihrer Erkennung zur Ausführung gebracht werden (Abb. 18).

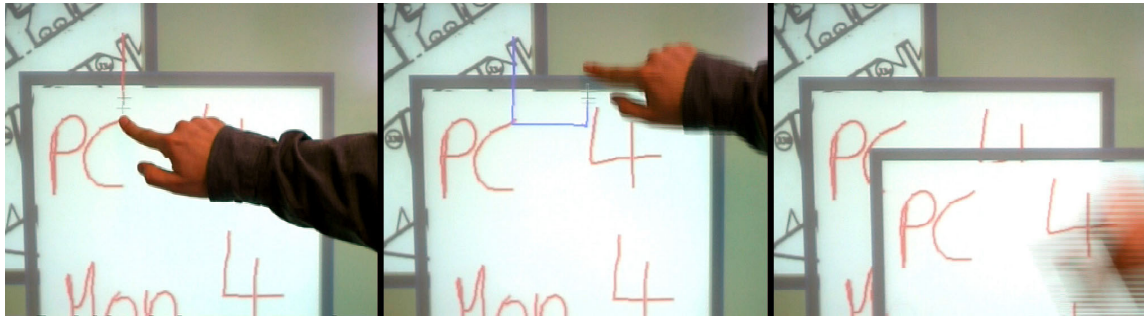


Abb. 18: Ausschnitte der Videosequenz einer Geste, mit der beispielsweise ein Objekt kopiert wird: kurz nach dem Beginn, kurz vor dem Ende der Interaktion und Ausführung der Geste (v.l.n.r.). Die Form der Geste ähnelt einem ‚U‘.

Für die Unterstützung dieser Interaktion wird der jeweils aktuelle Zustand des Prozesses der Gestenerkennung durch ein akustisches Signal repräsentiert, das aus dem Verlauf einer allgemeinen Interaktion hervorgeht. Dabei entsteht bei der Gesteninteraktion ein typisches zeitliches akustisches Muster, das als *Gestenmelodie* bezeichnet werden soll. Die Melodien aller Gesteninteraktionen beginnen mit ähnlichen akustischen Signalen, enden jedoch je nach Typ der erkannten Geste verschieden. Dabei steht weniger die Abfolge von Tonhöhen im Vordergrund, wie z. B. beim Ansatz der Earcons (Blattner et al. 1989, Brewster 1994), sondern die zeitliche Abfolge von akustischen Signalen, die eng mit dem Verlauf der Gestenerkennung verbunden sind. Die Geschwindigkeit beim ‚Spielen‘ der Gestenmelodie wird durch die Art der individuellen Interaktion bestimmt. Es handelt sich bei der Gesteninteraktion um eine höhere Form der akustischen Darstellung, die die realen Objekteigenschaften deutlich übertrifft. Die verschiedenen Gesten wie Erzeugen, Kopieren oder Löschen von Objekten unterscheiden sich durch einen spezifischen Abschlussklang (siehe Abb. 20). Die Syntax der Gestenmelodie ist in ihrer Struktur immer identisch. Die Semantik des Abschlussklangs bzw. -zeichens wird zur einfachen und deutlichen Unterscheidung aus elementaren Variationen des Grundgeräusches der Interaktion in Frequenz und Zeit gebildet.

In Abb. 17 ist beispielsweise der zeitliche Zusammenhang zwischen akustischer und optischer Darstellung detaillierter wiedergegeben. Die Werte für die Zeitangaben werden direkt durch die Interaktion gesteuert. Zum Zeitpunkt t_0 beginnt die Interaktion mit dem Aufsetzen des Stiftes oder Fingers, während t_3 das Ende der Interaktion darstellt, bei dem der Stift von der Oberfläche abgenommen wird. Die Zeitpunkte t_1 und t_2 werden während der kontinuierlichen Stiftbewegung aus dem Mechanismus der Gestenerkennung bestimmt und signalisieren die verschiedenen Zustände des Gestenerkennungsprozesses. Zum Zeitpunkt t_2 ist die Geste vom System erkannt und

der dazugehörige Befehl könnte schon ausgeführt werden. Jedoch erst, nachdem die Interaktion zum Zeitpunkt t_3 durch den Benutzer beendet wird, kommt es zur Ausführung des Gestenbefehls. Es kann zu Interaktionen kommen, bei denen nach dem Zeitpunkt t_2 eine erkannte, aber nicht intendierte Geste ignoriert wird. In diesem Fall wird nach dem Ende der Interaktion kein akustisches Signal gemäß der erkannten Geste ausgelöst. Man kann nach Nigay und Coutaz (1993) bei dieser parallelen und kombinierten Interaktion auch von synergetischer, multimodaler Interaktion sprechen (siehe dazu Abb. 3, Seite 16).

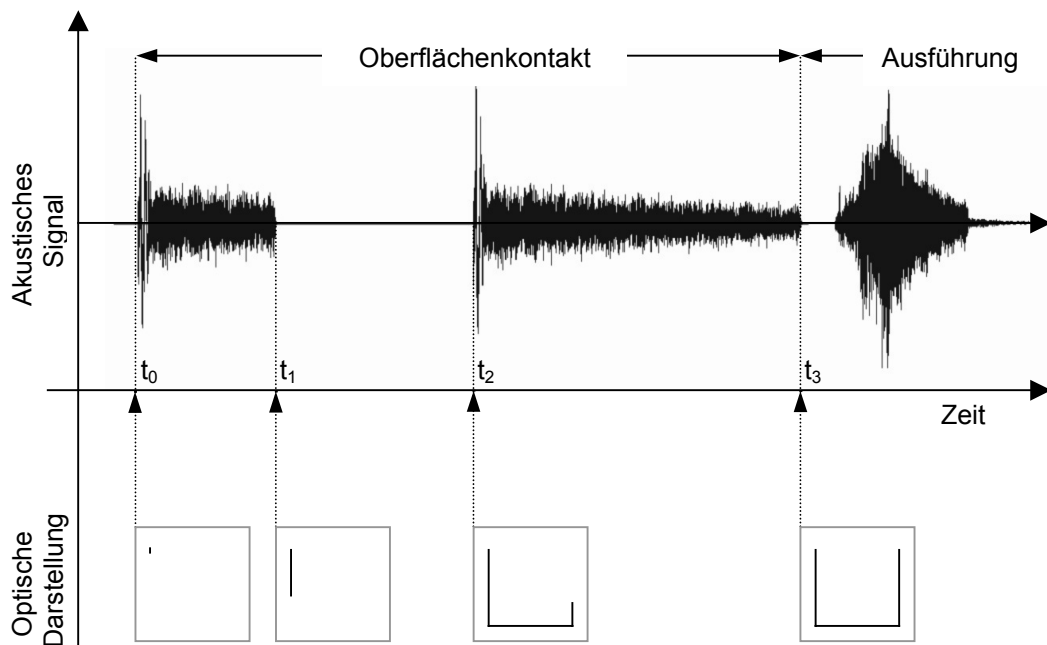


Abb. 17: Zusammenhang zwischen akustischer und optischer Darstellung im Verlauf einer Gesteninteraktion an der DynaWall.

Typische Gesteninteraktionen wie in Abb. 17 und Abb. 18 dargestellt, dauern bis zu einer Sekunde, so dass die Gestenmelodie eher als Einheit wahrgenommen wird, denn als Abfolge von einzelnen Teilsignalen. Im Gegensatz zu den perkussiven oder transienten akustischen Signalen aus dem täglichen Umfeld (wie z. B. bei Gaver 1993c beschrieben) wird die wichtigste Information bei den Gestenmelodien, d. h. der Typ der erkannten Geste, am Ende der *Melodie* vermittelt. Die Wirkung könnte wie folgt abgeschätzt werden: Der erste Teil der Gestenmelodie erzeugt Aufmerksamkeit und bewirkt eine Erwartungshaltung beim Benutzer. Das Abschlusssignal der erkannten Geste bestätigt dem Benutzer seinen intendierten Befehl.



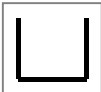

Geste	Abschlussklang der Gestenmelodie
Öffnen 	In der Frequenz aufsteigendes Bandrauschen, abklingend
Löschen 	In der Frequenz fallendes Bandrauschen, abklingend
Kopieren 	Wie „Öffnen“ nur mit einem leisen Echo
Abbrechen 	Tieffrequenter Sinuston für haptisch Darstellung

Abb. 20: Verschiedene Typen von Gesten und der jeweils zugehörige Abschlussklang der Gestenmelodie.

Mit einer besonderen Geste kann eine zuvor erkannte Geste abgebrochen bzw. der Befehl, der mit ihr verbunden ist, ignoriert werden. Für diese besondere ‚Abbrechen-Geste‘ (Abb. 20, unten) wurde mit einer speziellen akustischen Darstellung experimentiert, mit der die taktile Wahrnehmung des Benutzers angesprochen wird. Im unteren Grenzbereich der Wahrnehmung akustischer Signale geht die Empfindung der Schallwellen über eine taktile Wahrnehmung der mechanischen Bewegungen und Schwingungen der Oberfläche. Im Fall der haptischen Darstellung an der DynaWall wird die Projektions- und Darstellungsebene mit einem sehr tiefen Ton (ca. 59 Hz) so beschallt, dass das gesamte System in einen Zustand der Resonanz gebracht wird. Der Lautsprecher versorgt das System mit Schallenergie, so dass die Oberfläche, an der der Benutzer mit dem Finger oder Stift interagiert, in eine leichte Schwingung gerät, die taktil wahrnehmbar ist. Gleichzeitig kann man ein leises Summen und weitere Obertöne vernehmen, die im Resonanzsystem entstehen.

6.3.4 Interaktion und Kooperation mit der DynaWall

Die Merkmale der Unterstützung der Interaktion und Kooperation an der DynaWall durch die beschriebenen zusätzlichen akustischen Darstellungen betreffen sowohl die individuelle Interaktion an der DynaWall als auch Situationen, in denen Kooperation zwischen mehreren Personen stattfindet.

Die individuelle Interaktion wird durch die akustische Darstellung zu einer multimodalen Interaktion, bei der nach Robertson (1998) Interaktionsfehler durch Redundanz gemindert werden können. Durch die zusätzliche akustische Darstellung wird die Metapher der Tafel als konzeptuelles Angebot an den Benutzer im anwendungsunabhängigen Anteil der Benutzungsschnittstelle der DynaWall vervollständigt. Der Benutzer kann natürlicher und stärker in die Interaktion involviert werden. In den Fällen, in denen eine Wiederholung der gleichen Interaktion erfolgt, sollte berücksichtigt werden, dass die akustische Darstellung dazu neigt, lästig und störend zu wirken. Geeignete Maßnahmen können getroffen werden, um ähnlich wie bei Rücknahme der Lautstärke bei lang andauernden Stift-Annotationen einer Störung der Benutzer entgegenzuwirken.

Bei der Gruppenkooperation spielt die akustische Darstellung eine andere wichtige Rolle, wenn neben den Situationen, in denen zwischen den Gruppenmitgliedern Augenkontakt besteht, gewissermaßen für kurze Zeit eine Kooperation an einem Ort bzw. in einem Raum entsteht, bei der aber verteilt gearbeitet wird. Das oben beschriebene individuelle Feedback bei der Interaktion wird in der Gruppensituation zu einem akustischen Signal, das Informationen über den Aktivitätszustand von Personen vermitteln kann (*awareness*, siehe auch Beaudouin-Lafon 1994). Die akustische Darstellung kann dadurch Situationen unterstützen, ohne dass Blickkontakt zwischen den Gruppenmitgliedern bestehen muss. ‚Wirft‘ z. B. eine Person A von einer Seite der DynaWall der Person B an der anderen Seite ein Objekt zu, kann Person B das Eintreffen des Objektes durch dessen zusätzliche akustische Anreicherung antizipieren, noch bevor diese Person es sehen kann.

6.3.5 Erweiterung der technischen Umgebung der DynaWall

Für die Integration der akustischen Darstellung in die DynaWall musste die existierende Roomware-Komponente in zwei Bereichen erweitert werden. Auf der Seite der BEACH-Software mussten Programmanteile modifiziert und andere erweitert werden, um Informationen über die Interaktion und die Position von Objekten zu gewinnen. Dabei senden lokale Prozesse Nachrichten über die Interaktion mit den Objekten an einen zentralen Server, der das gemeinsame Datenmodell verwaltet und die Grundlage für die optische Darstellung der Objekte liefert. Die Daten der lokalen Prozesse steuern auch die Erzeugung der akustischen Signale (Sonifikation), die wiederum dem Benutzer über Lautsprecher zurückgeführt werden (siehe Abb. 20). Auf der anderen Seite war eine Erweiterung der Roomware-Komponente DynaWall notwendig: In die Rückprojektionseinheiten wurden die Lautsprecher integriert. Unter der Vorgabe der ‚Unsichtbarkeit‘ der Darstellungsmittel (Gestaltungsregel A) wurden die Lautsprecher

in die Rückprojektionseinheiten integriert. Dabei mussten Kompromisse hinsichtlich der Qualität der akustischen Darstellung eingegangen werden, die sich jedoch zugunsten der Wirksamkeit der Metapher der DynaWall als großflächige Tafel auswirken. Die zusammenhängende Wahrnehmung der optischen und akustischen Repräsentationen würde durch sichtbare Lautsprecher gestört, die sonst als Quelle des akustischen Signals identifiziert werden könnten und die Perzeption beeinflussen würden.

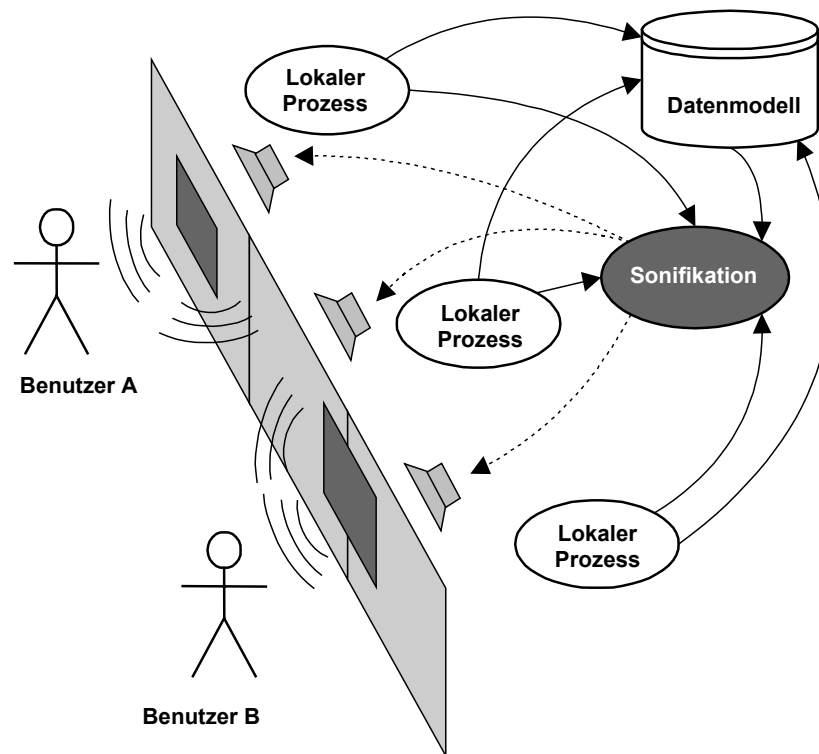


Abb. 20: Schematische Darstellung der verteilten Softwarekomponenten und die Erweiterungen zur Erzeugung und Darstellung von akustischen Signalen an der DynaWall.

6.3.5.1 Anbindung an die Softwareumgebung

Die DynaWall ist mit drei Computern ausgestattet, die über das Computernetzwerk mit einem Server verbunden sind. BEACH erzeugt mit dieser verteilten Anordnung über die Computerprojektionen der drei benachbarten Rückprojektionseinheiten eine räumlich zusammenhängende optische Darstellung an der DynaWall. An jedem der drei Segmente der DynaWall kann gleichzeitig jeweils eine Interaktion erfolgen. Innerhalb der Ereignisverwaltung von BEACH wurden Ergänzungen vorgenommen, die es

erlauben, die für die akustische Darstellung relevanten Informationen als Datenströme abzuleiten, damit die entsprechenden akustischen Signale erzeugt werden können. Zusätzlich wurden weitere Objektklassen gebildet, mit deren Hilfe über eingebundene Funktionsbibliotheken spezielle Hardware Schnittstellen, z. B. die MIDI-Ausgabe, kontrolliert und Daten ausgegeben werden können.

Im Gegensatz zur BEACH-Architektur, die aus einem zentralen Server und mehreren Clients besteht, ist die Erweiterung für die akustische Darstellung an der DynaWall ereignisgesteuert und besitzt keine zentrale Verwaltung eines Datenmodells (siehe Abb. 20). Die Informationen über die Interaktion werden in MIDI Kommandos entsprechend dem Typ der Interaktion kodiert und über jeweils eine direkte MIDI-Verbindung an den Audio-PC übertragen (Abb. 21). Die Entscheidung, anstelle der existierenden breitbandigen Netzwerkverbindung der Computersysteme die langsamere serielle Verbindung mit dem MIDI Protokoll zu wählen, liegt in den Eigenschaften der Verbindung begründet. Zwar können MIDI-Ereignisse nur mit einer Rate von 31.5 kBit/sec. zwischen zwei Geräten übertragen werden, jedoch ist diese Rate als Ressource gesichert und Verzögerungen unterliegen keiner Abhängigkeit z. B. von der Netzwerklast, wie es etwa bei Netzwerkverbindungen über das Internet-Protokoll der Fall sein kann (Best-Effort-Service).

Des Weiteren ist die Verarbeitung von MIDI-Ereignissen innerhalb des Computers mit geringer Verzögerung möglich, wie sie z. B. auch zur Ansteuerung von externen Klangsynthesizern in Echtzeit benutzt wird (Dannenberg 1996). Da jeder Client eine eigene MIDI-Leitung zum Audio-PC besitzt (siehe Abb. 21) wird vermieden, dass es bei mehreren gleichzeitigen Interaktionen zu Verzögerungen kommt. Vorrang vor einer Synchronisation mit der optischen Darstellung hat die minimale verzögerte Ausgabe von akustischen Signalen bei einer Interaktion (Gestaltungsregel D).

Die Alternative zum beschriebenen Ansatz der Ereignisverwaltung ist die Verwendung einer Model-View-Controller Architektur des Systems der Benutzungsschnittstellen in BEACH bzw. SmallTalk. Mit einem *acoustic view* wären damit systematisch auch verteilte Funktionalitäten bei der Ereignisverwaltung ermöglicht. Die verteilten Zustände und Interaktionen würden somit auf dem zentralen BEACH-Server in einem Datenmodell abgebildet werden können, aus dem dann die Erzeugung der Audiosignale erfolgen kann. Diesem Zugewinn an Möglichkeiten und Funktionalität durch die Ausnutzung der Eigenschaften der verteilten Programmierung in BEACH steht jedoch eine nicht zu tolerierende Schwankung der zeitlichen Verzögerung der akustischen Repräsentation entgegen. Sowohl die Synchronisation zur optischen Darstellung als auch die Verzögerung der Antwort auf eine Benutzereingabe waren deutlich als „out of

sync' Effekte wahrzunehmen, mit Werten deutlich oberhalb von 80 ms. Dies war insbesondere bei der initiierten automatischen Bewegung von Objekten und deren akustischer Darstellung immer dann festzustellen, wenn das Objekt im ‚fliegenden‘ Zustand die obere oder untere Grenze des Displays erreichte und dann gemäß der Reflexionsgesetze seine Richtung änderte. Das Auftreffen des Objekts an der Begrenzung der DynaWall wird durch ein zusätzliches akustisches Signal vermittelt. Der Zeitunterschied zwischen dem Einsatz des optischen und des akustischen Ereignisses war jedoch so groß und variierte so stark, dass *ein* kohärentes Ereignis dem Zuhörer nicht vermittelt werden konnte. Der Ansatz der zentralen Ereignisverwaltung wurde daraufhin nicht weiter verfolgt, da keine Möglichkeiten in Aussicht waren, die Synchronisation deutlich zu verbessern.

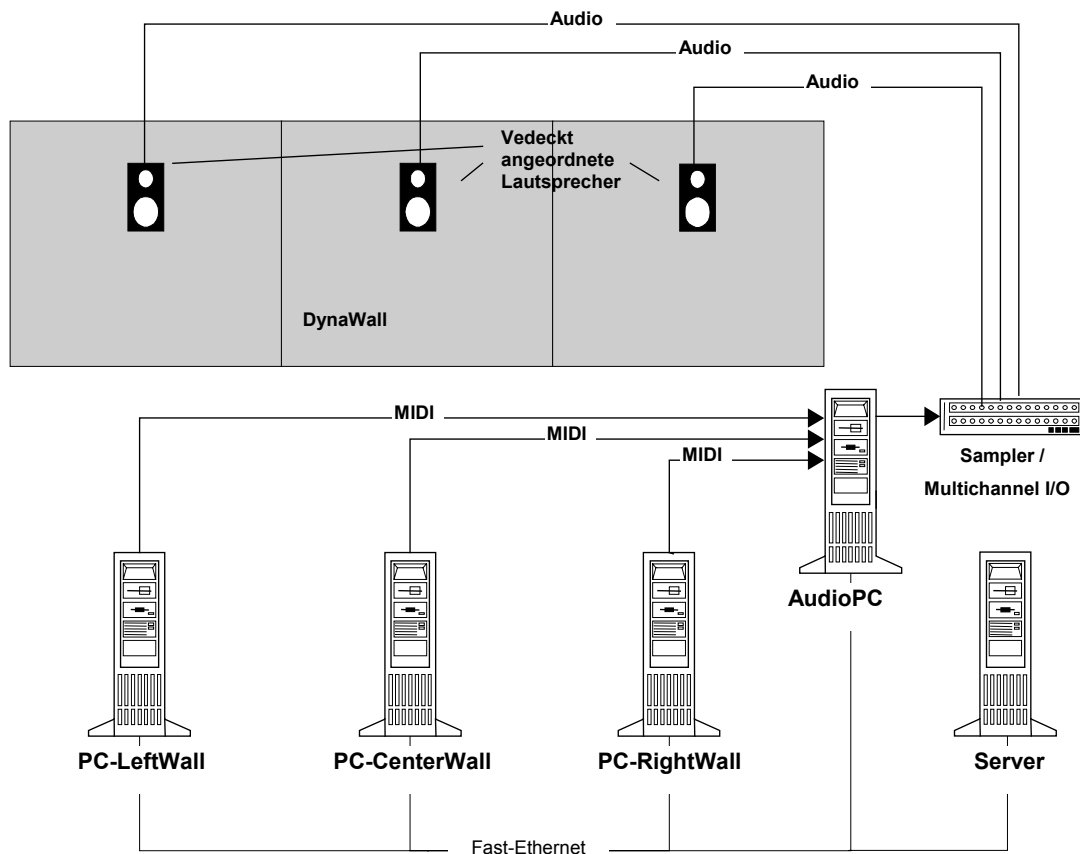


Abb. 21: Anordnung zur Realisierung der akustischen Repräsentation für die Interaktion mit Objekten an der DynaWall.

6.3.5.2 Anordnung zur Erzeugung der akustischen Darstellung

Die vorhandene Installation der DynaWall wurde um Komponenten ergänzt, die speziell für die Erzeugung und Generation von Audiosignalen verwendet wurden. Ein zentraler Audio-Computer (siehe Abb. 21) bildet das Bindeglied, das einerseits Nachrichten über die Interaktion an der DynaWall erhält und andererseits diese verarbeitet und daraufhin Audiosignale über eine mehrkanalige Einheit zur digital-analog Umsetzung an die Lautsprecher abgibt.

Die Aufgaben des Audio-Computers liegen in erster Linie darin, die eingehenden MIDI-Datenströme zu einem Hauptstrom zu vereinigen und an die Klanggeneration innerhalb des Computers weiterzuleiten. Die Klanggeneration wird durch einen frei konfigurierbaren Sample-Synthesizer übernommen, der auf einer PCI-Soundkarte vorhanden ist. Die Zuordnung der verschiedenen Klänge und Geräusche zu den Ereignissen der Interaktion erfolgt über das MIDI-Protokoll. Insofern verhält sich der Audio-Computer aus der Sicht der BEACH-Clients als festprogrammierter MIDI-Synthesizer (Müller-Tomfelde und Steiner 2001). In einer nächsten Stufe werden dann auf dem Audio-Computer die Audiosignale weiter verarbeitet, bevor sie umgesetzt und an das Lautsprechersystem abgegeben werden.

Bei der Interaktion mit Objekten auf der Oberfläche wirkt die Geschwindigkeit der Objekte auf ein Tiefpassfilter zur Klangmanipulation. Die Eckfrequenz des Tiefpassfilters wird in Echtzeit so gesteuert, dass schnelle Bewegungen ein breiteres Spektrum des Klangs hörbar machen. Zusätzlich wird die abnehmende Geschwindigkeit beim ‚Werfen‘ von Objekten durch eine abnehmende Resonanzüberhöhung an der Eckfrequenz des Tiefpassfilters dargestellt. Damit unterscheidet sich die akustische Darstellung der kontrollierten Positionierung deutlich von der Darstellung der ‚geworfenen‘ Objekte. Diese Funktionalitäten werden durch die Steuerung von Verarbeitungsroutinen bewirkt, die durch einen digitalen Signalprozessor bearbeitet werden, der sich auf der Soundkarte befindet. Die Routinen sind Bestandteil der Betriebssoftware der Soundkarte und können ebenso über MIDI kontrolliert werden.

Schließlich wird der Ort, an dem eine Interaktion an der DynaWall stattfindet, für die akustische Repräsentation wie folgt verwendet bzw. verarbeitet: Die horizontale Position der Interaktion eines Benutzers bestimmt das Lautstärkeverhältnis (Balance), mit dem das akustische Signal über die drei Lautsprecher der DynaWall ausgesendet wird. Dadurch erfolgt die akustische Repräsentation auch immer dort, wo Interaktion an der DynaWall durch den Benutzer stattfindet (Regel B). Die Balance wird auch bei den automatischen Bewegungsabläufen („Werfen“ von Objekten) dynamisch unterstützt.

Die Integration der Lautsprecher in die Rückprojektionseinheit erfolgte so, dass keine Abschattung der Darstellung der Computerprojektion durch den Lautsprecher auftreten kann (Regel A). Gleichzeitig sollte der Benutzer vor der Projektion auch im direkten Abstrahlungsfeld des Lautsprechers stehen, um die Oberfläche auch als ‚Quelle‘ der akustischen Signale wahrzunehmen. Für die Anforderungen wurde ein Ort in ca. 1.8 Metern Höhe und ca. 90 cm hinter der Projektionsfläche innerhalb der Projektionseinheit gefunden (siehe Abb. 22). In dieser Position ist die Abstrahlrichtung des Lautsprechers ca. 25 Grad nach unten geneigt und trifft in etwa die Mitte der Rückprojektionsfläche. Bei der Wiedergabe mit einer solchen Anordnung ergeben sich unerwünschte Resonanz- und Dämpfungseffekte, d. h. Verzerrungen des Signals. Das abgestrahlte akustische Signal wird im Wesentlichen durch die Rückprojektionsscheibe in den hohen Frequenzen gedämpft und zusätzlich wird die Richtcharakteristik des Lautsprechers stark verändert. Teilweise wurden diese Einflüsse durch geeignete Vorverzerrungen kompensiert und die Schwingungsresonanzen der Rückprojektionseinheiten durch Absorptionsmaterialien reduziert.

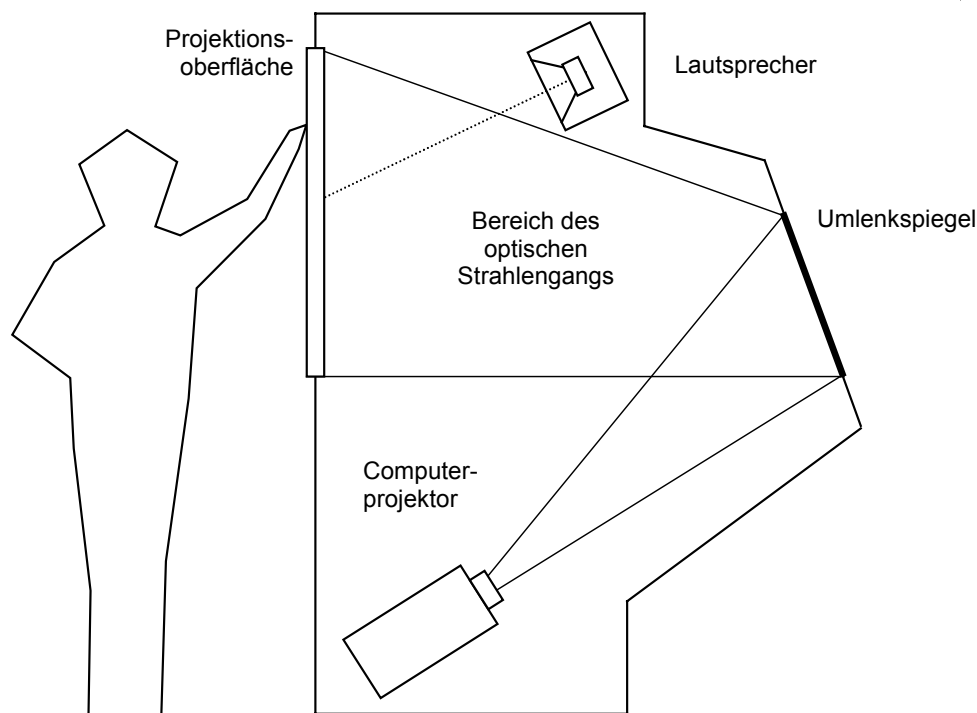


Abb. 22: Seitenansicht einer Rückprojektionseinheit der DynaWall mit der Kombination der Einrichtungen zur akustischen und optischen Darstellung. Der optische Strahlengang der Computerprojektion wird durch den Lautsprecher nicht beeinträchtigt.

6.4 **Erweiterte akustische Modellierung von Oberflächeninteraktionen**

Das Konzept der virtuellen Reibung, wie es an der DynaWall realisiert ist, bietet weitere Möglichkeiten der akustischen Darstellung bei der Interaktion über die Vermittlung der Geschwindigkeit von Objekten hinaus. So können je nach Konstellation von Objekt und Oberfläche verschiedene Reibungsverhältnisse bestehen, die sich in Abhängigkeit von speziellen Anforderungen bei der Interaktion auch über der Zeit ändern können. Aus der Anwendung heraus können aber auch die akustischen Eigenschaften der Oberfläche in Abhängigkeit vom Kontext, in dem sich der Benutzer in der Anwendung befindet, gesteuert werden. Der Versuch, auf der interaktiven Oberfläche in einem Bereich zu schreiben, in dem dies zum Zeitpunkt der Interaktion im Kontext der Anwendung nicht erlaubt ist, kann durch entsprechende akustische Signale dargestellt werden. Das Geräusch des Schreibens auf einer Glasoberfläche könnte z. B. vermitteln, dass die ‚Tinte‘ in dem Bereich nicht ‚haften‘ bleibt und so daran erinnern, dass in diesem Bereich der Oberfläche nicht geschrieben werden kann.

Eine weitere Motivation für die Entwicklung dieser indirekten Darstellung der Oberflächeneigenschaften durch eine akustische Darstellung liegt darin begründet, dass ein solches Verfahren sehr einfach auf eine Vielzahl von berührungsempfindlichen Oberflächen von Computerdarstellungen ohne zusätzlichen Aufwand angewendet werden kann. Diese akustischen Darstellungen fungieren als Zeichen (Index) der Objekteigenschaften, die durch die fehlende haptische Darstellung jedoch unvollständig an der Systemoberfläche wiedergegeben werden. Auf Effekte synästhetischer Wahrnehmung (engl. cross-modal) kann in diesem Zusammenhang verwiesen werden, wonach die Stimulation *einer* Modalität (z. B. des Hörsinns) dazu führen kann, dass eine *andere* Modalität mit stimuliert wird (wie z. B. der Tastsinn).

Aus der Perspektive der ökologischen Psychologie (siehe dazu Abschnitt 3.2.1) nimmt ein Hörer über die Geräusche Eigenschaften von Objekten auf Oberflächen als Invarianzen wahr. Er kann z. B. auf die Materialien schließen, die an einem Prozess beteiligt sind (Gaver 1993c). Bei der in Abb. 23 dargestellten Situation werden z. B. typische Geräusche bei der Interaktion erzeugt, aus denen auf die Art des verwendeten Stiftes geschlossen werden kann. Durch den Prozess der Interaktion werden die strukturellen Invarianzen der Objekte wie Material oder Oberflächenbeschaffenheit erkennbar bzw. hörbar und vermitteln dem Benutzer Handlungsmöglichkeiten. Im Gegensatz zu optischen Darstellungen, bei denen die strukturellen Invarianzen sofort sichtbar sind, werden die Invarianzen in akustischen Darstellungen erst bei der

Wahrnehmung über einen Zeitraum hörbar. Gleichzeitig werden jedoch bei der Interaktion mit den Objekten auch Transformations-Invarianzen, d. h. typische Merkmale zeitlicher Abläufe, vermittelt wie z. B. die Art oder Geschwindigkeit, mit der ein Objekt bewegt wird, unabhängig davon, aus welchem Material es besteht. Mit zunehmender räumlicher Ausprägung der Oberfläche können ihre Invarianzen dann auch visuell wahrnehmbar werden, wie z. B. im Falle einer Holzmaserung, bei der die anisotrope Textur der Oberfläche sichtbar ist.

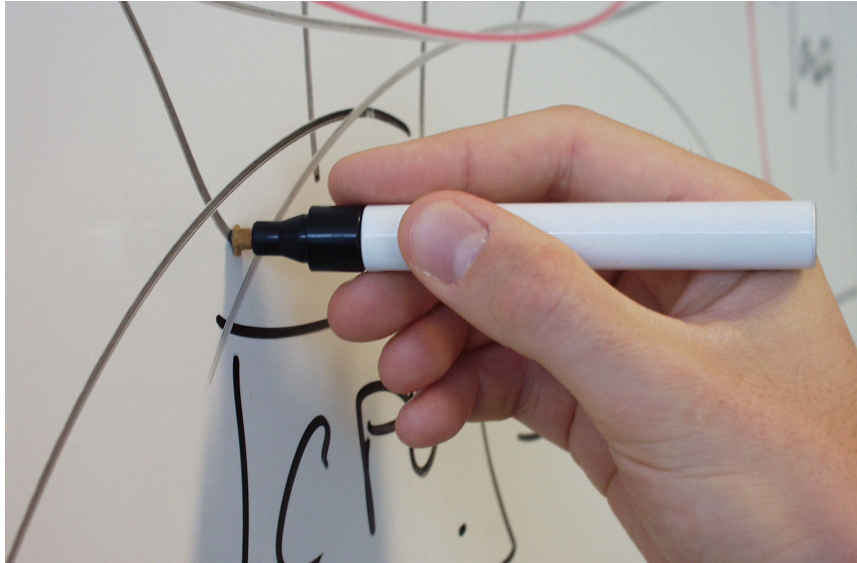


Abb. 23: Beispiel einer typischen Stiftinteraktion an einer Tafel mit einem sog. Boardmarker.

6.4.1 Modell der Stiftinteraktion

Das physikalische Modell, das zu einem einfachen Synthesemodell führen soll, ist in Abb. 24 dargestellt. Die Oberfläche, hier eine Tischplatte, wird durch die Spitze eines zylindrischen Stiftes berührt und der Kontaktpunkt wird bei der Interaktion über die zweidimensionale Oberfläche bewegt. Aus den verschiedenen Ableitungen des Orts nach der Zeit lassen sich verschiedene Informationen über die Interaktion gewinnen. In einem ersten Schritt erhält man den Betrag der Geschwindigkeit, mit der der Stift bewegt wird, und die Richtung der Interaktion aus der Ableitung des Ortes nach der Zeit. In einem zweiten Schritt bzw. durch Ableitungen höherer Ordnung kann dann noch eine Aussage über die Charakteristik der Interaktion getroffen werden, z. B. ob es sich um eine Translation oder Rotation handelt. Besteht während der Interaktion keine Änderung der Richtung, kann die Interaktion als Translation bezeichnet werden. Ist die Änderung der Richtung jedoch über der Zeit konstant, handelt es sich um eine Rotation.

Je größer der Wert dieser konstanten Änderung ist, desto kleiner ist der Kreis, der bei der Interaktion vollzogen wird.

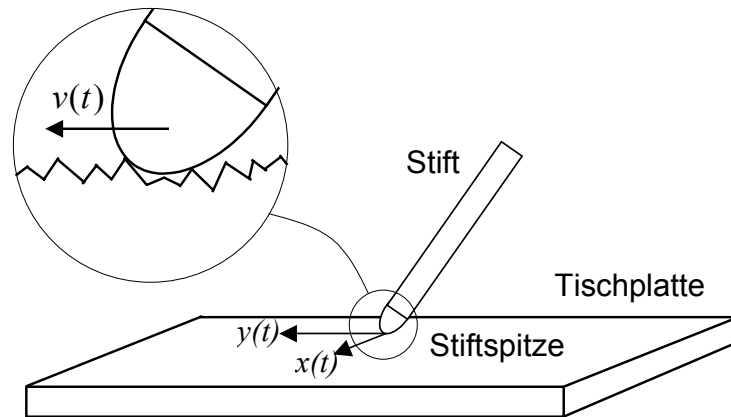


Abb. 24: Das einfache physikalische Modell der Interaktion mit einem Stift auf einer Oberfläche. Die Stiftspitze wird auf der zweidimensionalen Oberfläche mit der Geschwindigkeit $v(t)$ bewegt. Kollisionen der Stiftspitze mit kleinen Hervorhebungen der Oberfläche erregen das akustische System.

Ein weiterer Teil des Modells in Abb. 24 ist in der Detailansicht hervorgehoben. Die Stiftspitze gleitet über die Oberfläche und je nach Beschaffenheit kommt die Spitze des Stiftes mit kleinen Erhebungen und anderen Unregelmäßigkeiten in Kontakt und führt zu einer Abfolge von Kollisionen, welche die Erzeugung der akustischen Signale stimuliert (Gaver 1993b). In einer ersten Näherung werden die Einflüsse von Druck und Beschleunigung bei der Interaktion für das Modell vernachlässigt. Vergleichbar ist dieses Modell mit dem einer Abtastnadel von Schallplattenspielern, bei denen ein Stereo-Signal in einer Spur auf der Oberfläche einer Schallplatte abgetastet wird (van Doel et al. 2001). Im Gegensatz zum Modell der Abtastnadel ist die Oberfläche in zwei Dimensionen ausgeprägt, wobei die Tischplatte und der Stift die Resonanzkörper der Anordnung darstellen, über die der Schall u. a. an die Umgebung abgegeben wird.

6.4.2 Modell passiver akustischer Objekte

Die zwei wichtigsten akustischen Objekte, die bei der Abstrahlung der Schallwellen bei der Interaktion eine Rolle spielen, sind der Stift und das Volumen ‚unter‘ der Interaktionsoberfläche. Der Stift wird als Zylinder und das Volumen als Rechteck modelliert (siehe Abb. 24). Die akustischen Eigenschaften solcher elementaren Volumen sind gut erforscht (Dutilleux und Müller-Tomfelde 1999, Rocchesso 2001) und weil die Formen mathematisch einfach zu beschreiben sind, sollen sie als *Basic Shapes* bezeichnet werden. Im Gegensatz dazu werden die in der natürlichen Umgebung vorkommenden Volumen als komplex bezeichnet (*Complex Shapes*). Die spektralen

Eigenschaften der *Basic Shapes* können durch Modelle angenähert bzw. imitiert werden, ohne dass aufwändige Verfahren zur Berechnung der akustischen Eigenschaften verwendet werden müssen, wie z. B. bei der Vorherbestimmung von Raumakustiken (Kleiner et al. 1993). Derartige Modelle beschränken sich auf die Imitation der charakteristischen akustischen Eigenschaften (also der strukturellen Invarianzen), liefern aber keine Aussagen über die konkrete Schallausbreitung und -abstrahlung im Raum über der Zeit. Für einen rechteckigen Resonator können mit der Gleichung

$$f_r(l, m, n) = \frac{c}{2} \sqrt{\left(\frac{l}{X}\right)^2 + \left(\frac{m}{Y}\right)^2 + \left(\frac{n}{Z}\right)^2} \quad (4)$$

die Frequenzen f_r bestimmt werden, die als Resonanzen für die Form des Volumens typisch sind (Zölzer 1997). Die Angaben X, Y und Z bezeichnen die Längen der verschiedenen Seiten des rechteckigen Volumens. Die Indizes l, m und n geben die jeweilige Ordnung der sog. Mode an. Die Ausbreitungsgeschwindigkeit c ist in der Regel abhängig vom Material, in dem sich die Schallwellen ausbreiten. Im Fall des zylindrischen Modells für den Stift lautet die Gleichung für die Grundfrequenz der Resonanz

$$f_r^c(l) = \frac{c}{2} \frac{l}{X}, \quad (5)$$

wobei X die Länge des Zylinders bzw. Rohrs bedeutet. Der Durchmesser des Rohrs kann dazu verwendet werden, die Frequenzabhängigkeit des Reflexionskoeffizienten zu imitieren (Dutilleux und Müller-Tomfelde 1999), hat aber keinen Einfluss auf den Wert der Resonanzfrequenz des Rohrs.

Für beide Modelle existieren einfache und effiziente Algorithmen der digitalen Signalverarbeitung, die hauptsächlich aus komplex klingenden Filterstrukturen wie sog. Kamm- und Tiefpassfiltern bestehen. Diese linearen und zeitinvarianten Filtersysteme (linear time-invariant, LTI) sind erforscht und in ihren akustischen Eigenschaften vorhersehbar bzw. ‚vorherhörbar‘. Das günstige Verhältnis zwischen nötiger Verarbeitungsleistung in Echtzeit und der komplexen Wirkung derartiger Filter auf das Spektrum eines Signals liegt in der Rekursivität der Filterstrukturen (infinite impulse response, IIR) begründet und wird deshalb auch oft zur Erzeugung von Raumnachhall verwendet (Schröder 1970, Roads 1996).

Ziel der Modellierung der verschiedenen Volumen ist es, eine einfache Kontrolle über die unterschiedlichen Anteile im gesamten akustischen Signal zu haben. Für die

Interaktion können dann z. B. die ‚Gegenstände‘, mit denen interagiert wird, die dieselben bleiben, während sich die Qualität der Oberfläche ändert. Ähnlich dem Ansatz von Beaudouin-Lafon (1994) können dann Kombinationen von verschiedenen Tischvolumen und -materialien, der Art des Stiftes und der Oberfläche erstellt werden.

6.4.3 Stimulation des akustischen Systems

Während die passiven akustischen Objekte eine Art starren Resonanzkörper darstellen, ist die Stimulation des akustischen Systems durch die Bewegung des Stiftes auf der Oberfläche ein dynamischer und zeitvarianter Prozess. Der wichtigste Parameter, der die Stimulation des Systems kontrolliert, ist die Geschwindigkeit der Stiftspitze auf der Oberfläche. Vergleicht man diese Interaktion mit dem Spielen auf einer Geige, fallen folgende Parallelen auf: Die Geschwindigkeit ist der Interaktionsparameter, mit dem ‚Energie‘ in das System geführt wird. Der Resonanzkörper der Geige entspricht den oben beschriebenen passiven Objekten wie Tischplatte und Stift bei der Stiftinteraktion. Die eigentliche Stimulation erfolgt dort, wo die lineare Bewegungsenergie in Schwingungsenergie umgewandelt wird, d. h. dort, wo der Bogen auf der Saite ansetzt bzw. der Stift die Oberfläche berührt (Farina et al. 1995).

6.4.4 Akustische Texturen von Oberflächen

Anstatt eines harmonischen Tons beim Streichen der Saite einer Geige werden bei der Interaktion mit Oberflächen Rauschprozesse stimuliert, die verschiedene Qualitäten von Oberflächen vermitteln. Die verbale Beschreibung dieser Qualitäten mit Begriffen wie rau, grob, weich oder glatt sind hauptsächlich entlehnt aus dem Bereich taktiler oder visueller Wahrnehmung. Im Folgenden werden drei verschiedenen Kategorien akustischer Darstellungen für Oberflächeneigenschaften und Texturen beschrieben (Mikro-, Meso- und Makro-Texturen), die die verschiedenen Möglichkeiten in Anwendungen mit Stiftinteraktion aufzeigen.

Mikro-Oberflächentextur:

Bei dieser Kategorie einer Oberflächentextur wird eine Eigenschaft über der Ebene als konstant und homogen vorausgesetzt. Die einzelnen Erhebungen der Oberfläche sind so verteilt, dass keine Direktionalität erkennbar ist. Weder auditiv noch visuell oder taktil können einzelne Anteile der Oberfläche identifiziert oder in der Ebene exakt lokalisiert werden. Die Größe der Oberflächenerhebungen ist stochastisch verteilt und diese Eigenschaft wird auf das akustische Signal übertragen, wenn die Stiftspitze über die Oberfläche bewegt wird. Die räumliche Anordnung von kleinen Erhebungen mit verschiedenen gaußverteilten Größen wird bei konstanter Bewegung der ‚Abtastspitze‘

über die Lateralkraft dieser Reibungskonstellation in eine zeitliche Abfolge von Kollisionen transformiert. Die Varianz und die Färbung des dadurch entstehenden Rauschprozesses sind unabhängig von der Geschwindigkeit. Varianz und spektrale Färbung stellen invariante akustische Eigenschaften des Materials dar (Siira und Pai 1996, van Doel et al. 2001).

Meso-Oberflächentextur:

Im Gegensatz zur vorhergehenden Kategorie soll es sich bei der Meso-Oberflächentextur um einzelne wahrnehmbare Erhebungen handeln wie etwa bei Holzmaserungen oder bei grobem Sandpapier. Die Textur dieser Kategorie kann eine Struktur haben, die in der Ebene eine Orientierung besitzt, so dass sich in Abhängigkeit von der Interaktionsrichtung die Stimulation ändert. Die Verteilung dieser Textur in der Ebene der Interaktionsoberfläche soll hauptsächlich stochastisch sein.

Makro-Oberflächentextur:

Diese Kategorie ist eng verbunden mit den optischen Eigenschaften der Oberfläche wie z. B. bei einer Struktur eines Schachbretts, bei der sich Bereiche mit verschiedenen Eigenschaften abwechseln. Auch können räumliche Begrenzungen von Oberflächenelementen können auch dazu gezählt werden. Diese Textur-Kategorie kann außerdem verstanden werden als Anordnung von Bereichen, in denen verschiedene Mikro- und Meso-Texturen vorherrschen.

Die drei Kategorien der Oberflächentexturen schließen sich in ihrer Existenz gegenseitig nicht aus, sondern vielmehr kann bei der Synthese einer Textur von der Möglichkeit der Superposition Gebrauch gemacht werden. Dadurch entsteht eine Vielzahl von Kombinationen, die zur Gestaltung akustischer Oberflächeneigenschaften verwendet werden können. Wie schon angedeutet können durch Mikro-Texturen strukturelle akustische Invarianzen vermittelt werden, die nicht sichtbar sind. Mit den Meso- und Makro-Texturen wird die Möglichkeit zusätzlicher Invarianzen geschaffen, die auch zum Teil visuell wahrgenommen werden können. So wird die Art der Textur bei einer kreisenden Interaktion mit einem Stift auf einer gerichteten Meso-Textur durch zyklische Veränderungen in der akustischen Darstellung wahrnehmbar.

6.4.5 Geräuschgenerator

Die Erzeugung der akustischen Signale durch Stimulation an der Spitze des Stiftes wird durch die Überlagerung von Einflüssen der Mikro- und Meso-Texturen gebildet. Der zentrale Interaktionsparameter für die Signalerzeugung ist die Geschwindigkeit, mit der der Stift über die Oberfläche bewegt wird: Er steuert eine Vielzahl von weiteren Parametern, die die Eigenschaften der Oberfläche bestimmen. In Abb. 25 sind diese

Eingabesignale des Generators links dargestellt und bestehen hauptsächlich aus vier Werten. Alle vier Eingabesignale (Abb. 25, rechts) werden in ihren Werten durch gaußsche Rauschprozesse moduliert, die in der Varianz eingestellt werden können.

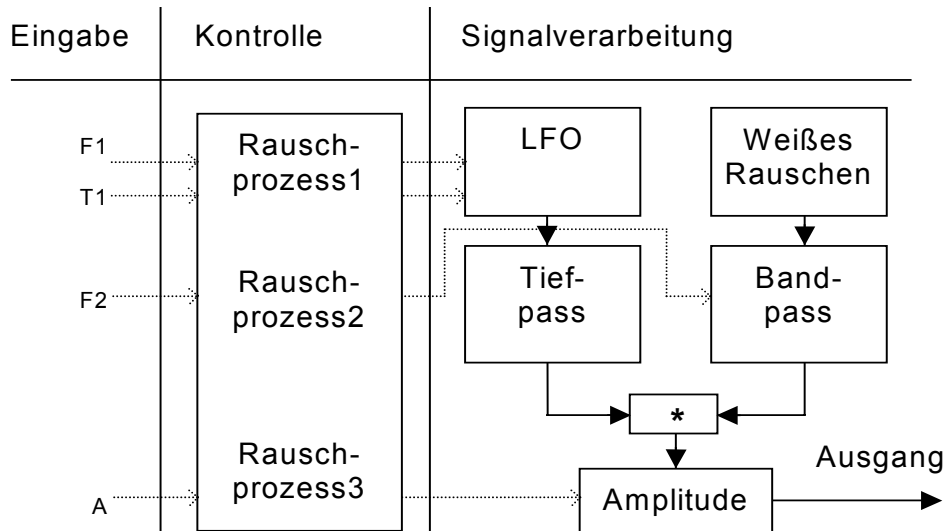


Abb. 25: Die schematische Darstellung des Geräuschgenerators: Ein Oszillator mit einer niedrigen Frequenz (LFO) moduliert das Signal eines Rauschgenerators. Die Eingabeparameter kontrollieren indirekt die Signalverarbeitung. Ihren Werten werden stochastische Charakteristika durch Rauschprozesse zugefügt.

Die Kontrollsignale des Generators bekommen dadurch einen stochastischen Charakter. Den Kern bildet ein Generator für weißes Rauschen, der durch ein einfaches Bandpassfilter in seinen spektralen Eigenschaften kontrolliert werden kann (rechts in Abb. 25). Damit wird die Mikro-Textur modelliert, wobei die Mittenfrequenz des Bandfilters auf der Eingabeseite als F2 Parameter verfügbar ist. Dieser Teil des Generators ist unabhängig von der Geschwindigkeit der Interaktion.

Die Eigenschaften der Meso-Textur werden durch den linken Teil im Signalbereich des Generators modelliert (siehe Abb. 25). Ein Oszillator für einen niedrigen Frequenzbereich (low-frequency-oscillator, LFO) liefert ein Rechteck-Signal, dass durch die Parameter F1 und T1 in seiner Frequenz und Pulsweite kontrolliert wird. Die Rechteckimpulse des Signals repräsentieren die kleinen Kollisionen der Stiftspitze mit Partikeln auf der Oberfläche. Die Unregelmäßigkeit einer Meso-Textur kann z. B. über die Varianz des Rauschprozesses von F1, der Frequenz des LFO, kontrolliert werden. Diese Frequenz muss sich proportional zum zentralen Interaktionsparameter bzw. zur Geschwindigkeit des Stiftes verhalten. Langsame Stiftinteraktionen bewirken ebenso eine langsame Abfolge von kleinen Kollisionen und somit eine geringe Frequenz F1, während rasche Interaktionen eine schnelle Abfolge der Kollisionen und daher eine

große Frequenz F1 bedingen. Dieser Zusammenhang muss optimal abgestimmt sein, da sonst die Eigenschaften der Meso-Textur nicht plausibel vermittelt werden. Mit T1 wird die Pulsweite des Rechtecksignals kontrolliert, wobei dieser Parameter die Größe der Erhebungen auf der Oberfläche repräsentiert. Im weiteren Verlauf glättet ein Tiefpassfilter die Flanken des Rechtecksignals, bevor es mit dem Rauschprozess der Mikro-Textur multipliziert wird. Die Filterung durch den Tiefpass kann dazu verwendet werden, um weniger ‚harte‘ Kollisionen zu imitieren, da der Einsatz des Signals weniger abrupt erfolgt. Die Gesamtlautstärke des Geräuschgenerators wird schließlich durch den Parameter A bestimmt. Neben der Geschwindigkeit kann hier auch die Druckstärke bei der Stiftinteraktion einen Einfluss auf die Gesamtlautstärke haben, wenn diese durch das Eingabegerät zu Verfügung steht. Oberflächeneigenschaften einer Mikro-Textur werden hauptsächlich im Frequenzbereich des Signals abgebildet, während die Eigenschaften einer Meso-Textur sich durch besondere zeitliche Verhältnisse auszeichnen.

6.4.6 Testumgebung für akustische Oberflächentexturen

Auf der Grundlage von physikalischen Betrachtungen über die Stiftinteraktion und Analysen realer Aufnahmen wurden in der ersten Phase der Gestaltung in einer Max/ MSP (Max Signal Processing) Umgebung einfache, akustische Modelle realisiert. Die grafische Programmierung in Max erlaubt den schnellen Aufbau von Prototypen (Puckette 1991a und 1991b, Max/ MSP 1997). Die Ergebnisse der Synthese können sehr schnell überprüft und iterativ verbessert werden (siehe Abb. 11, Seite 49). Zur Ansteuerung des Generators wurden in Max die Koordinaten der Computermaus abgefragt und damit die Geschwindigkeit der Interaktion mit einem Stift errechnet. Ein prinzipielles Problem der Kontrollsignale tritt dabei in Erscheinung: Die Abtastrate der Ortskoordinaten der Interaktion in Computerumgebungen liegt deutlich unter der von digitalen Audiosignalen und kann, wie an verschiedenen Eingabegeräten gemessen, von 10Hz bis ca. 100Hz reichen. Dies führt besonders bei kurzen und schnellen Eingaben zu Problemen durch Aliasing-Effekte und damit zu Irritationen beim Benutzer. Die Folge ist eine geringe Plausibilität der Oberflächeneigenschaften bei der Interaktion (siehe dazu Abschnitt 4.2.2).

In einer zweiten Phase wurde der Prototyp des Generators in einer Windows-Umgebung aufgebaut und mit einem Grafiktablett als Eingabegerät versehen. Der Generator konnte dadurch weiter entwickelt werden und mit dem Eingabegerät wurden zusätzliche Parameter der Interaktion verfügbar wie Druck, Anstellwinkel und Orientierung des Stiftes (Müller-Tomfelde und Münch 2001).

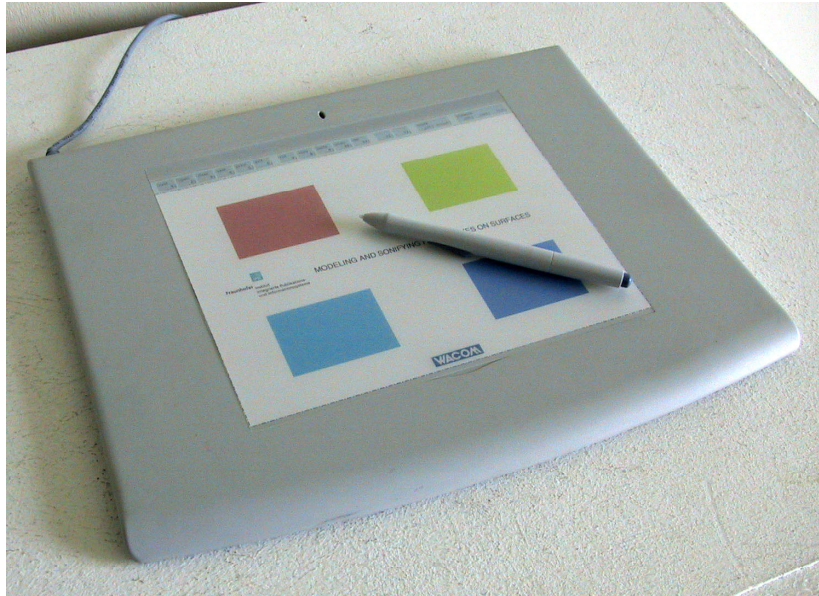


Abb. 26: Einfache Anordnung zur Demonstration der verschiedenen Kategorien von Oberflächentexturen. Bei der Interaktion mit dem Stift auf den verschiedenfarbigen Flächen werden unterschiedliche Textur-Kombinationen hörbar.

In einer einfachen Anwendung wurden alle drei Kategorien von Texturen entwickelt und erprobt (siehe Abb. 26). Die vier verschiedenfarbigen Flächen repräsentieren verschiedene Kombinationen von Mikro- und Meso-Texturen auf der Oberfläche, während der weiße Hintergrund lediglich eine uniforme Mikro-Textur besitzt. Weitere Parameter der Interaktion können gewonnen werden, die das Modell der Stiftinteraktion erweitern und zusätzliche akustische Effekte erlauben. Aus dem statischen Anstellwinkel des Stiftes in der Ebene und der Richtung der Bewegung der Stiftspitze (Orientierung) kann bestimmt werden, ob der Stift über die Oberfläche ‚gezogen‘ oder ‚geschoben‘ wird. Im letzteren Fall neigt sich der Stift entgegen seiner Bewegungsrichtung. Wird dieses ‚Schieben‘ akustisch leicht unterschiedlich zum ‚Ziehen‘ dargestellt, werden ansonsten unverändert klingende Interaktionen, wie z. B. bei einer kreisenden Stiftbewegung durch zyklische Klangveränderungen, hörbar und dienen als Hinweis auf die Art der Interaktion.

6.5 Passage – Physikalisches ‚Lesezeichen‘ in die virtuelle Welt

Der Passage-Mechanismus wurde entwickelt, um einen einfachen und intuitiven Transport von Objekten zwischen verschiedenen Orten in hybriden Umgebungen zu ermöglichen (Konomi et al. 1999). Die Metapher des Übertritts von der realen in die virtuelle Umgebung und vice versa soll dem Benutzer schon durch den Begriff *Passage* (Durchgang, Durchfahrt, Überfahrt) nahe gelegt werden: Auf der ‚Überfahrt‘ befindet

sich ein sog. *passenger*, der an einem speziellen Ort über eine Brücke zwischen der realen und virtuellen Umgebung wechseln kann (siehe Abb. 27).

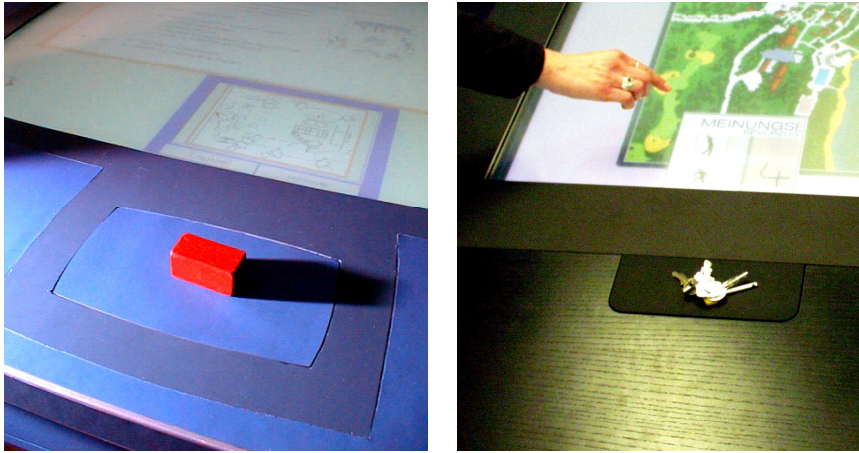


Abb. 27: Eine ‚Bridge‘ des Passage-Mechanismus besteht aus einem realen und einem virtuellen Teil (Bereich der Computerdarstellung). Dem roten Klötzchen (links) oder dem Schlüsselbund (rechts) ist als ‚passenger‘ jeweils ein Objekt in der Computerdarstellung zugeordnet (hier am InteracTable).

Im Gegensatz zur Übertragung von Daten zwischen zwei Computern z. B. mittels Disketten wird beim Passage-Mechanismus eine Referenz zwischen einem realen physikalischen Objekt und einem Objekt innerhalb der Computerumgebung etabliert. Mit dieser Referenz wird von verschiedenen Orten der Zugriff mittels des realen physikalischen Objekts auf ein Objekt innerhalb der Computerumgebung ermöglicht. Dieser neuartige Transportmechanismus erlaubt den intuitiven Umgang mit Datenobjekten in hybriden Umgebungen. Im Gegensatz zu herkömmlichen und abstrakten Methoden der Übertragung über das Computernetzwerk erfolgt der Transport über reale, greifbare Gegenstände aus der Umgebung. Die Interaktion des Benutzers besteht aus der Abfolge von drei wesentlichen Operationen: die Herstellung der Referenz, d. h. Zuordnung eines Objektes in der Computerdarstellung zum realen Objekt, gefolgt vom Transport des realen Objektes zu einem anderen Ort und schließlich dem Abrufen des Inhalts über das reale Objekt an einer sog. Bridge.

6.5.1 Anwendung und Einordnung des Passage-Mechanismus

Konzipiert wurde der Mechanismus für solche Situationen kooperativen Arbeitens, in denen herkömmliche Verfahren des Datentransports von einem Ort zum anderen die Kooperation blockieren. Mit herkömmlichen Verfahren ist z. B. das Abspeichern auf einem Medium wie Disketten etc. oder das Zwischenspeichern auf einem zentralen Fileserver gemeint. Zum einen werden vom Benutzer explizite Eingaben in

verschiedener Form und Abfolge benötigt, um die Daten in ein Transportmedium zu überführen (und wieder zurück) und zum anderen muss eine Erinnerungsleistung während des ‚Transports‘ erbracht werden, um z. B. den Ort oder Namen einer temporären Datei in Erinnerung zu behalten. Der Passage-Mechanismus ermöglicht mit einer einfachen Interaktion die Überführung in das ‚Transportmedium‘, das gleichzeitig als reales Objekt eine Art ‚materialisierte Erinnerung‘ darstellt.

Nach dem Zeichenmodell für Mensch-Computer Interaktionen nach Familant und Detweiler (1993) kann man beim Passage-Mechanismus auch von einer *indirekten Referenz* sprechen, bei der die Eigenschaften des Zeichenreferenten (passenger) in der Regel keine Übereinstimmungen mit dem denotativen Referenten teilen. Aus semiotischer Sicht spricht man dann auch von einer symbolischen Repräsentation. Die Besonderheit des Passage-Mechanismus liegt nun darin, dass der Zeichenreferent nicht nur eine optische Darstellung besitzt, sondern auch physisch berührbar ist, wobei die Bildung der Referenz zum Gegenstand der Interaktion wird. In einer Übersicht sog. *Tangible Interface Instances* (Ullmer und Ishii 2000) wird der Passage-Mechanismus daher auch als „assoziatives System“ eingeordnet. Während des Prozesses der Herstellung der Referenz bzw. des Abrufens der Referenz erhält der Benutzer auf der Signal- und Interaktionsebene direkten Zugang zur Referenz: Auf der einen Seite der ‚Bridge‘ liegt das reale Objekt, während auf der anderen Seite in der Computerdarstellung ein korrespondierende Objekt sichtbar wird (siehe auch Abb. 27). Mit direkter Manipulation kann der Benutzer dem realen Objekt ein Element der Computerdarstellung zuordnen und damit im Sinne der Passage-Metapher die ‚Brücke überschreiten‘, d. h. die Referenz herstellen. Das spezifische Objekt in der Computerumgebung geht über in eine physische, symbolische Repräsentation. Nach dem Transport im realen Raum wird auf gleicher Weise der ‚transportierte Inhalt‘ wieder durch eine einfache Interaktion auf der Computeroberfläche verfügbar. Der Passage-Mechanismus kann daher durch seine Eigenschaft, realen Gegenständen zusätzliche Bedeutung hinzuzugeben, in den Bereich der *Augmented Reality* (siehe auch Abb. 4, Seite 19) eingeordnet werden. Der einfache und intuitive Umgang mit Informationen sowohl innerhalb der Computerumgebung als auch in der realen Umgebung an verschiedenen Orten ist aber auch als ein Beispiel aus dem Bereich *Ubiquitous Computing* zu verstehen (siehe Abschnitt 2.3.2).

6.5.2 Architektur für den Passage-Mechanismus

Damit der Passage-Mechanismus an verschiedenen Orten ermöglicht werden kann, ist die zugrundeliegende Architektur des Systems im Raum verteilt und besitzt einen zentralen Server zur Verwaltung der Daten. Der Passage-Mechanismus verläuft beim

Aufruf wie folgt: Nachdem das reale Objekt auf der Bridge identifiziert wurde, erfolgt über einen Webserver eine Anfrage an eine Datenbank. Mit der Angabe der Identität des Objektes wird aus der Datenbank das zugehörige Objekt ausgelesen und über den Passage-Agenten an die Applikation weitergegeben, die das Objekt als virtuellen Teil der Bridge anzeigt (siehe Abb. 28).

Zur Identifikation der realen Objekte, die auf die Bridge gelegt werden, sind zwei Technologien mit verschiedenen Eigenschaften in Verwendung:

- eine Präzisionswaage mit serieller Kommunikationsschnittstelle und
- die Identifikation mithilfe sog. Radio-Frequency-Identification (RFID)-Tags.

Mit der Präzisionswaage ist es möglich, einen beliebigen Gegenstand, der auf die Bridge (Waage) gelegt wird, über sein Gewicht zu identifizieren. Die kleinste Gewichtseinheit, die aufgelöst werden kann, beträgt ca. 0.2 g, so dass in einem Gewichtsbereich von 0 bis ca. 50 g ca. 250 Objekte mit unterschiedlichen Gewichtswerten identifiziert werden können. Dabei kann es zu uneindeutigen Situationen kommen, wenn z. B. zwei unterschiedliche Objekte den selben Gewichtswert haben, da sie dann nur *ein* Objekt in der Computerumgebung repräsentieren können. Bei zwei Interaktionen an unterschiedlichen Bridges kann dies zu Konfliktsituationen führen. Eine zusätzliche Modellierung der räumlichen Anordnung der Bridges könnte in solchen Situationen Informationen liefern, mit denen zwischen zwei Interaktionen mit verschiedenen Objekten mit gleichem Gewicht unterschieden werden kann. Minimale Transportzeiten von Bridge A nach Bridge B bilden dann die Grundlage, um mit erhöhter Wahrscheinlichkeit zwischen Objekten mit selben Gewicht zu differenzieren.

Wird die Identifikation über RFID-Tags realisiert, existiert dieses Problem bei der Identifikation nicht. Dafür muss jedoch das zu identifizierende Objekt speziell präpariert sein. Das Prinzip der Identifikation besteht darin, dass eine batterielose elektronische Schaltung mit einer eindeutigen Kennung (Tag, siehe Abb. 29) in das elektromagnetische Erregerfeld (125 kHz) des Auslesegeräts (Lesegerät, Reader, siehe Abb. 29) gebracht wird, so dass die elektronische Schaltung dadurch Energie aufnehmen kann. Danach sendet die elektronische Schaltung ihre eindeutige Kennung (32Bit-Schlüssel) an den Reader. Das Prinzip der Datenübertragung wird als Lastmodulation bezeichnet (Finkenzeller 2002). Diese RFID-Technologie wird häufig für die Realisierung von Szenarien im Kontext von *Ubiquitous Computing* (siehe Abschnitt 2.3.2) verwendet.

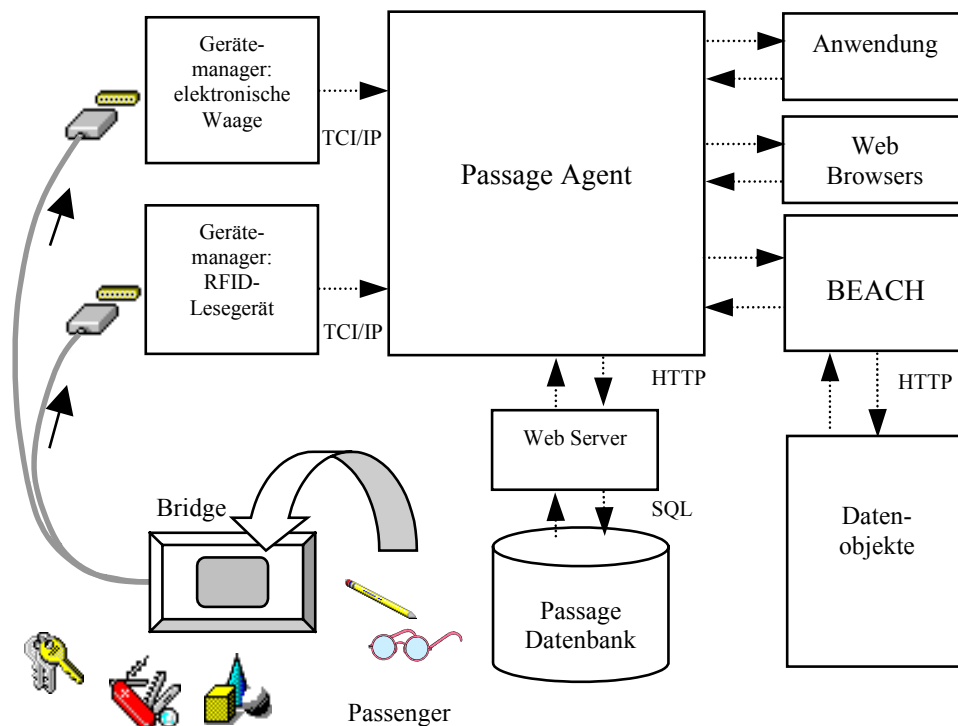


Abb. 28: Die verteilte Architektur für den Passage-Mechanismus: Der Passage-Agent verwaltet die Kommunikation zwischen der Bridge (linke Seite) und den Anwendungen, die den Mechanismus benutzen (rechte Seite).

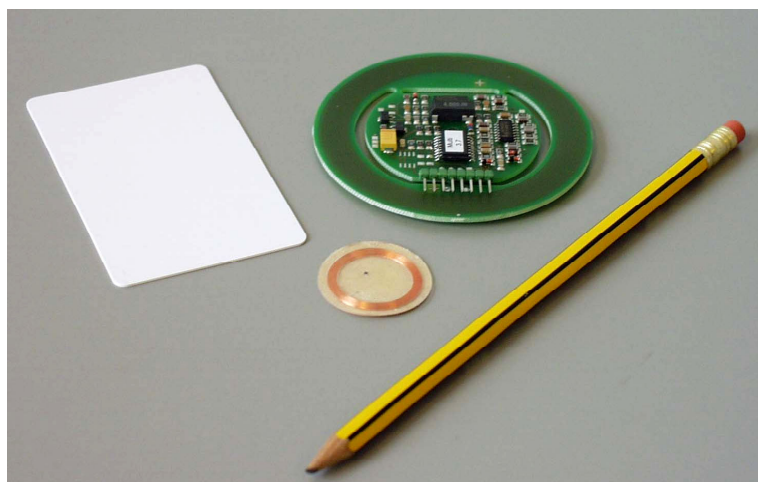


Abb. 29: Die Elemente der Identifikation mit der Transponder-Technologie: Die sog. Reader-Elektronik (rechts oben), ein Transponder als aufklebbare Miniaturschaltung (Mitte) oder als weiße Kunststoffkarte mit integrierter elektronischer Schaltung (links; Bleistift dient zum Größenvergleich).

Der Hauptunterschied der beiden Identifikationstechniken liegt darin, dass mit der Waage beliebige Objekte für den Passage-Mechanismus verwendet werden, während

mit den präparierten Objekten eine eindeutige Identifikation erfolgt und verlässlich sowie langfristig z. B. persönliche Daten damit verbunden werden können.

6.5.3 Unterstützung der Interaktion durch akustisches Feedback

Der Transportmechanismus mit realen Objekten ist intuitiv und schnell verständlich, wohingegen der Prozess der Herstellung der Referenz als abstrakt zu bezeichnen ist. Um diese Ungleichheit im konzeptuellen Modell auszugleichen, soll dem Benutzer mit einer akustischen Repräsentation der Prozess der Interaktion besser vermittelt werden. Die typische Interaktion beim Passage-Mechanismus kann in drei nacheinander abfolgenden Teilinteraktionen beschrieben werden, wobei der mittlere Teil abhängig davon ist, in welcher Richtung die Brücke des Passage-Mechanismus ‚begangen‘ wird. Die drei Teilinteraktionen sind:

- das Auflegen eines Passengers bzw. reales Objektes auf die Bridge,
- das Herstellen einer Referenz zwischen dem Passenger und einem Objekt der Computerdarstellung bzw. das Erzeugen eines Objekts der Computerdarstellung aufgrund einer bereits etablierten Referenz und
- das Abnehmen des Passengers.

Technisch bedingt entstehen Zeitverzögerungen, die aus der Ansprechzeit des Identifikationsverfahrens, der Netzwerkübertragung der Datenbankanfrage und des Datentransfers zusammengesetzt sind. Signifikant ist z. B. die Verzögerung durch die elektronische Waage, die auf das Messverfahren mit einem Dehnungsmessstreifen zurückzuführen ist. Der Gewichtswert des Objektes auf der Waage wird über der Zeit durch die Messelektronik asymptotisch angenähert, bis schließlich der stabile Messwert ausgegeben werden kann. In der Regel kann dieser Vorgang bis zu drei Sekunden dauern. Um diesen Zeitraum nicht ohne weitere Signalisierungen auf der Systemoberfläche für den Benutzer verstreichen zu lassen, wurde ein zusätzliches akustisches Feedback realisiert, um Irritationen und Unsicherheiten bei der Benutzung zu begegnen.

Der Passage-Mechanismus erzeugt keine realen Geräusche, womit mögliche akustische Repräsentationen frei wählbar sind. Die Grundlage der akustischen Repräsentation bildet lediglich der zeitliche Ablauf des Passage-Mechanismus. Dabei vermittelt die akustische Darstellung symbolisch das Verhalten bzw. den Status, in dem sich der Prozess der Identifikation und Zuordnung befindet. Für den Zeitbereich des Auffindens des assoziierten Objektes in der Datenbank (Suchen) und zum Zeitpunkt der

Identifikation und Verfügbarkeit des Objektes auf dem virtuellen Bereich der Bridge (Gefunden) wurden verschiedene Schemata in der akustischen Darstellung realisiert.

So wird dem Benutzer mit einem musikalischen Schema während der Identifikations- und Zuordnungsphase ein schnelles Arpeggio⁴ zugespielt, das die Suche nach dem Datenobjekt repräsentieren soll. Wurde das Objekt in der Datenbank gefunden und auf dem virtuellen Teil der Bridge angezeigt, erklingt dann synchron zur Erscheinung der optischen Darstellung im Gegensatz zum Arpeggio ein Akkord. Die Elemente der akustischen Repräsentation sind einzelne Töne einer Gitarre, deren zeitlich verschiedene Anordnung unterschiedliche Zustände vermittelt. In Anlehnung an die Versuche von Warren und Verbrugge (1984) soll durch diese Repräsentation die besondere Fähigkeit der menschlichen auditiven Wahrnehmung zur Entdeckung von zeitlichen Mustern ausgenutzt werden.

Für eine abstrakte akustische Darstellung bei der Interaktion wird der Zeitraum des Suchens mit einem tiefen, pulsierenden Ton überbrückt, der durch leise, kurze und bandbegrenzte Rauschabschnitte begleitet wird, die in ihrer Mittenfrequenz zufällig variieren. Dieser Klang geht in einen einzelnen Ton über, der aus dem tiefen Ton entsteht und sich in seiner Grundfrequenz zu höheren Frequenzen bewegt, wenn das assoziierte Datenobjekt gefunden und angezeigt wird.

Die optischen und akustischen Darstellungen bei dieser Interaktion sprechen nicht nur unterschiedliche Modalitäten der Wahrnehmung an, sondern berücksichtigen die verschiedenen Ausprägungen der visuellen und auditiven Perzeption und lenken dadurch die Aufmerksamkeit des Benutzers im Verlauf der Interaktion. Nachdem der Benutzer den Passenger auf die Bridge gelegt hat, wartet er auf die Antwort des Systems, um seine Interaktion fortzusetzen. In dieser Zeit wird die Aufmerksamkeit des Benutzers schwach durch das periphere Feedback der akustischen Darstellung aufrechterhalten (Arpeggio bzw. tieffrequenter, pulsierender Ton), das den Benutzer an anderen Interaktionen wie z. B. einer Kommunikation mit Blickkontakt nicht hindert. Am Ende des Suchvorgangs fordern die oben beschriebenen Klänge eine größere Aufmerksamkeit des Benutzers und lenken sie auf die fokussierte optische Darstellung der Bridge.

⁴ Spielart eines Musikinstruments, bei der die Töne eines Akkords nacheinander zum Erklingen gebracht werden.

Die Plausibilität der gesamten Interaktion wird durch eine zusätzliche analoge Korrespondenz zwischen dem Objekt der Interaktion (Passenger) und der akustischen Darstellung des Interaktionsprozesses erhöht. Der Wert des Gewichts des Passengers, wird im Verlauf seiner Bestimmung durch die Wägezelle dem wahren Gewichtswert asymptotisch angenähert. Dieser Verlauf der Bestimmung des Gewichts wird als Parameter zur Beeinflussung der akustischen Darstellung verfügbar und z. B. mit der Lautstärke verknüpft (Gestaltungsregel C aus Abschnitt 6.2.), so dass leichte, reale Objekte ein leiseres Feedback erhalten als schwere.

Der Passage-Mechanismus mit der zusätzlichen akustischen Darstellung ist an den Roomware-Komponenten DynaWall und InteracTable realisiert. Die Anordnung der Lautsprecher bei der letzteren Roomware-Komponente folgt auch der Gestaltungsregel B aus Abschnitt 6.2, wonach die akustische Darstellung dort erfolgen sollte, wo sie für die Interaktion relevant ist. Unterhalb der beiden Bridges am InteracTable (siehe Abb. 15, rechts, Seite 59) ist jeweils ein Lautsprecher installiert, der die oben beschriebenen Systemzustände beim Passage-Mechanismus akustisch wiedergibt.

6.6 Akustische Darstellung von Zuständen in Computersystemen

Die Erfahrungen mit den Roomware-Komponenten und insbesondere mit dem CommChair (Müller-Tomfelde und Reischl 1998) und ConnecTable (Tandler et al. 2001) haben gezeigt, dass durch die Integration der Computer in die Elemente des Raums ihre physischen Eigenschaften, die ihre Erscheinung ausmachen, immer weiter in den Hintergrund treten (*disappearing computer*). Technologische Fortschritte wie z. B. die Miniaturisierung von Systemkomponenten (Festplatte etc.) führen dazu, dass der Computer als Gerät kleiner wird und auch seine akustischen Eigenschaften verschwinden. Am Beispiel der Roomware-Komponente ConnecTable soll daher untersucht werden, was eine Forderung nach hörbaren Objekteigenschaften (siehe Abschnitt 5.2) in diesem Zusammenhang bedeuten und wie sie erfüllt werden kann.

6.6.1 Der geräuschlose Computer

Der ConnecTable ist als ein Büromöbel konzipiert, das ähnlich wie ein Schreibpult verwendet wird. Durch seine Mechanik kann der ConnecTable sowohl im Sitzen als auch im Stehen verwendet werden. An der Stelle der Schreibfläche befindet sich ein Computerdisplay, das mithilfe eines speziellen Stiftes bedient werden kann. Ein vollständiger Computer samt Energieversorgung befindet sich im Anbau der Standsäule des ConnecTables (siehe Abb. 30). Bei der Konzeption des Computersystems wurden Komponenten verwendet, die keine Geräusche verursachen. Die Kühlung des

Prozessors erfolgt passiv über einen Kühlkörper ohne zusätzlichen Ventilator und als Festplatte wurde ein Typ gewählt, der aus statischen Speicherbausteinen aufgebaut ist (Tandler et al. 2001).



Abb. 30: Der ConneCTable ist ein Schreibpult, das in seiner Höhe justiert werden kann. Die Komponenten der Computertechnik befinden sich im Anbau an die Standsäule des Möbelstücks. Die Schreibfläche besteht aus einem Computerdisplay, das mit einem Stift bedient werden kann.

Die Konsequenz dieser Entwicklung ist ein Computersystem, das nicht nur als geräuscharm, sondern vielmehr als geräuschlos bezeichnet werden kann. Beispiele von Produkten aus dem Computerbereich, die ähnliche besondere akustische Eigenschaften besitzen, sind der iCube von Apple, der eine vollständige passive Kühlung des Versorgungsnetzteils besitzt oder Speicherfestplatten von IBM, bei denen das Geräusch beim Datenzugriff eingestellt und an die Bedürfnisse der Verwendung angepasst werden kann (IBM Feature Tool 2001).

Im Gegensatz zu den Geräuschen bei der direkten Interaktion, wie sie für die Stiftinteraktion beschrieben wurden (siehe Abschnitt 6.4), sollen nun Geräusche im Vordergrund stehen, die als ‚Reaktion‘ des Computersystems als Gerät auf die Einwirkung durch den Benutzer entstehen (siehe dazu Abschnitt 5.3). Auf eine explizite Interaktion des Benutzers antwortet der Computer mit einer Abfolge von internen Prozessen, die der Benutzer über akustische Signale vermittelt bekommt. Ein typisches Beispiel hierfür ist die Operation des Abspeicherns eines Dokuments in einer

Anwendung. Der Benutzer drückt die Tastenkombination ‚Ctrl-s‘ auf der Computertastatur und das Computersystem ‚reagiert‘ mit einer wahrnehmbaren Veränderung des Geräuschs der Festplatte, auf der die Daten der Anwendung gespeichert werden. In einem anderen Beispiel wird bei der Interaktion mit einem Laptop indirekt eine erhöhte Prozessorbeltastung durch das Einsetzen des Geräuschs vom Lüftungsgebläse erkennbar, wenn der Computer eine aufwändige Berechnung für den Benutzer ausführt.

6.6.2 Der akustische Systemmonitor

Um die Forderung nach hörbaren Objekteigenschaften bei der Interaktion mit dem ConneCTable zu erfüllen, wurde eine Umgebung zum Experimentieren mit verschiedenen akustischen Darstellungen geschaffen. Die Umgebung ermöglicht den Zugriff auf die Parameter der Systemleistung, mit denen akustische Darstellungen kontrolliert werden können. Im Unterschied zum Ansatz von Cohen (1994a und 1994b) liegt der Fokus dieses akustischen Systemmonitors nicht auf der Darstellung abstrakter Prozesse, wie sie bei verteilten Zugriffen auf die lokale Festplatte des Computers entstehen. Es sollen die internen Zustände des Computers akustisch dargestellt werden, um die Eigenschaft des Computers als Gerät wieder herzustellen und neuartige Qualitäten hinzuzufügen, die im Anwendungskontext nützlich sein können. Dabei können auch Situationen entstehen, die keinen realen Bezug mehr besitzen, wenn z. B. eine Datei auf der Festplatte eines zentralen Servers gespeichert wird und das akustische Feedback dafür jedoch lokal erfolgt.

Unter Verwendung einer Schnittstelle zum Windows-Betriebssystem (Performance Data Helper Interface, PDH) können durch sog. Counter die Zustände des Systems wie z. B. benutzte Prozessorzeit ausgelesen werden. Die Aktivität einer Festplatte beim Lesen und beim Schreiben kann ermittelt werden sowie die Anzahl der Daten pro Zeiteinheit, die in Paketen über die Netzwerkschnittstelle in und aus dem Computer gelangen. Für einen Prototyp wurden diese drei Parameter verwendet, um exemplarisch einen Klanggenerator anzusteuern.

Die akustische Darstellung der Systemzustände erfolgt bei dem hier beschriebenen Ansatz im Hintergrund einer jeden Anwendung und parallel zu anderen akustischen Signalen bei der Interaktion. So können z. B. auch gleichzeitig die sog. akustischen System- und Programmsignale der grafischen Benutzungsschnittstelle von Windows erklingen. Die akustische Darstellung der Systemzustände wird nicht mit einer speziellen Anwendung in Verbindung gebracht, sondern gehört zum Computer als Gerät. Die akustische Darstellung vermittelt eine Präsenz des Computers als

physikalisches Gerät unabhängig von seiner speziellen Verwendung auch dann, wenn der Computer physisch gar nicht mehr sichtbar ist.

Andere Ansätze und Formen der Darstellung von Informationen aus der Computerumgebung sind z. B. in der Installation von Natalie Jeremijenko 1995 am Xerox Palo Alto Research Center realisiert worden (Weiser 1996). Der sog. *Dangling String* ist eine Plastikschnur, die an einem elektrischen Schrittmotor an die Decke montiert ist. Jedes Datenpaket, das über das lokale Netzwerk übertragen wird, gibt einen Drehimpuls auf den Motor, wodurch starke Netzwerkaktivität durch das Schwirren des Kabels hörbar wird. Weitere Beispiele akustischer Darstellungen von Parametern von Computerumgebungen sind im sog. *AmbientROOM* realisiert (Ishii und Ullmer 1997), in dem z. B. ein einzelner Aufruf von speziellen Webseiten eines Webserver den Klang eines niederprasselnden Regentropfens auslöst. In diesen beiden Ansätzen werden die Daten der Computerumgebung direkt und abstrakt dargestellt und das technisch vermittelnde Gerät, wie z. B. der Webserver, wird nicht weiter berücksichtigt. Der akustische Systemmonitor für geräuschlose Computer soll aber gerade die Existenz des Computers als Gerät vermitteln und damit an das mentale Modell des Benutzers anknüpfen.

6.6.3 Klanggenerator für Systemaktivitäten

Für die Erzeugung der Klänge für die Systemzustände eines Computers wurde ein abstraktes akustisches Modell gebildet, das nicht aus einer reinen additiven Synthese besteht, sondern die Systemzustände in *ein* komplexes Klangobjekt abbildet. Es sollte nicht versucht werden, die Computergeräusche zu imitieren, sondern vielmehr die Systemaktivität künstlich und dennoch plausibel darzustellen. Das dazu verwendete Modell besteht aus einem modulierten Generator, der Rauschimpulse produziert, ähnlich wie in der akustischen Modellierung der Meso-Texturen (siehe Abschnitt 6.4.4). Die Frequenz des Generators wird durch die Gesamtlast der Prozesse kontrolliert, die auf dem Computer zur Ausführung kommen. In einer weiteren Verarbeitungsstufe wird dieses Signal, das eine feste Resonanzfrequenz besitzt und in der Güte der Resonanz kontrolliert wird, einer Kammfilterung unterzogen. Mit dem Kammfilter wird dem Ausgangssignal des Generators je nach Aktivität an der Netzwerkschnittstelle mehr oder weniger stark eine harmonische Eigenschaft aufgeprägt, die der Benutzer im Gesamtsignal wahrnehmen kann.

Die Prozessorauslastung als Grundgeräusch der Gesamtdarstellung verändert sich hauptsächlich in der zeitlichen Struktur des Klangs. Geringe Prozessaktivität auf dem Computer stellt sich als langsames An- und Abschwellen eines Rauschsignals dar. Je

mehr die Last auf dem Prozessor zunimmt, um so schneller wird der Rhythmus der Rauschimpulse, bis das Geräusch in ein sog. Knattern übergeht, das Ähnlichkeit mit Motorgeräuschen besitzt. Über die Güte der nachfolgenden Kammfilterstruktur wird die Netzwerkaktivität abgebildet. So nimmt der Benutzer bei keiner oder geringer Netzwerkaktivität lediglich das oben beschriebene akustische Verhalten wahr. Nimmt die Netzwerkaktivität hingegen zu, wird die Grundfrequenz des Filters immer deutlicher hörbar, bis ein harmonischer Ton im Rhythmus des Generators deutlich hörbar wird. Der Einfluss dieses Parameters wirkt sich ausschließlich im Frequenzbereich aus, die Deutlichkeit der Gestalt des Tons, d. h. die ‚Kante‘ des auditiven Objektes (siehe Abschnitt 3.1, Seite 26) korrespondiert mit der Netzwerkaktivität des Computers.

Im Hinblick auf die Art, wie die akustische Darstellung der Aktivitätszustände erfolgt, kann man nach Kramer (1994) auch von einer sog. Sonifikation sprechen. Die Daten einer Quelle sind nicht direkt hörbar, sondern steuern die Erzeugung akustischer Signale. Vom Standpunkt der Mensch-Computer Interaktion ist diese akustische Darstellung unabhängig von der speziellen Anwendung und vermittelt die ‚technische Reaktion‘ des Computers als Gerät (akustisches Verhalten im anwendungs-unabhängigen Teil der Benutzungsschnittstelle, Abschnitt 5.3).

Neben der abstrakten und symbolischen akustischen Repräsentation der internen Zustände des Computers wurde auch mit einer metaphorischen Repräsentation experimentiert. Dazu wurde die Metapher des Plattenspielers für die Repräsentation verwendet (Müller-Tomfelde 2002b). Es wird dabei für die Anwendungssituation davon ausgegangen, dass der Benutzer während der Interaktion dem Abspielen einer Musikdatei zuhört. Je mehr die Prozessorleistung des Computersystems durch eine Anwendung des Benutzers belegt wird, desto langsamer wird eine Musikdatei wiedergegeben. Der Effekt kommt dem ‚Abbremsen‘ einer Schallplatte auf dem Plattenteller gleich und vermittelt dem Benutzer, dass das System ‚belastet‘ ist. Wird zusätzlich noch auf die Festplatte des Systems zugegriffen, werden dem abgespielten Klang kleine Artefakte zugefügt, die ähnlich wie Kratzer auf alten Schallplatten klingen. Damit wird über die Analogie zwischen der rotierenden Schallplatte und der Festplatte auf die Aktivität verwiesen, dass auf Daten der Festplatte zugegriffen wird.

6.7 Hybride Präsentation akustischer Signale

Neben den verschiedenen Möglichkeiten der Verwendung von akustischen Signalen bei der Interaktion soll auch auf die Form der Präsentation eingegangen werden. Gegenstand der Betrachtungen soll nicht das sein, *was* dargestellt ist, sondern es soll untersucht werden *wie* ein akustisches Signal präsentiert wird und welche Rolle die

charakteristischen Eigenschaften der Präsentation für die Mensch-Computer Interaktion spielen. Auch wenn die Verwendung des Begriffs Darstellung sowohl für die visuelle als auch auditive Wahrnehmung verwendet wird, unterscheiden sich ihre Eigenschaften so deutlich (Gaver 1997), dass Aussagen, die für optische Darstellungen gelten, nicht ohne Weiteres auf akustische Darstellungen übertragen werden können. So ist das übliche Mittel zur optischen Darstellung in Computerumgebungen in Form eines Standardbildschirms für mehrere Personen gleichzeitig nicht ausreichend, wohingegen akustische Darstellungen mit Lautsprechern von solchen Einschränkungen befreit und im ganzen Raum wahrnehmbar sind. Im Folgenden soll die hybride Präsentation vorgestellt werden, mit deren Hilfe z. B. Anwendungen für *Augmented Reality* auch im akustischen Medium realisiert werden können.

6.7.1 Kontinuum der Präsentation akustischer Darstellungen

Um die Bedeutung der verschiedenen Arten der Präsentation von akustischen Signalen einordnen zu können, soll in Anlehnung an die Taxonomie optischer Darstellungen von Milgram und Kishino (1994) ein Kontinuum der Präsentationsformen akustischer Darstellungen erstellt werden (siehe auch Abschnitt 2.3.1). Dabei können Analogien z. B. zwischen den Techniken optischer Darstellung im Bereich der *Augmented Reality* zu den Techniken der akustischen Darstellung gezogen werden. Mit sog. *see-through displays*, wie sie im KARMA System (Knowledge-based Augmented Reality for Maintenance Assistance, von Feiner et al. 1993) verwendet werden, wird eine computergenerierte Darstellung über die tatsächliche Ansicht des Benutzers projiziert und die ‚Realität‘ erweitert. Analog dazu kann man bei akustischen Darstellungen von einem *hear-through display* sprechen, wenn den akustischen Signalen aus der Umwelt eine künstliche akustische Darstellung überlagert wird. Für die Darstellung von akustischen Signalen stehen im Prinzip zwei Techniken zur Verfügung, die bezüglich der Bedeutung für die auditive Wahrnehmung, zumindest in statischen Hörsituationen, ähnlich sind. Es handelt sich dabei um die Präsentation mit Lautsprechern und die mit Kopfhörern. Für unterschiedliche Anwendungskontexte besitzen beide Verfahren der Präsentation jedoch sehr verschiedene und teilweise gegensätzliche Eigenschaften.

Auf dem Kontinuum der *Mixed Auditory Reality* befinden sich links in Abb. 31 einerseits reale Schallquellen und rechts andererseits Formen der Präsentation, bei der der Zuhörer von der realen akustischen Umgebung getrennt ist und vollständig virtuelle Schallquellen wahrnimmt. Die letztere Form der Darstellung erfolgt z. B. mit geschlossenen Kopfhörern (siehe Abb. 31). Einen Schritt weiter als reale Schallquellen in Richtung rein virtueller akustischer Umgebungen ist die Präsentation mithilfe von Lautsprechern. Mit mehreren Lautsprechern ist es möglich, sog. Phantomschallquellen

zwischen den Lautsprechern zu platzieren. In der einfachsten Form geschieht dies über den Intensitätsunterschied zwischen zwei Lautsprechern (siehe Abschnitt 3.1.3, Seite 29). In komplexen Anwendungen, wie z. B. in virtuellen Umgebungen (Cruz-Neira et al. 1992; Eckel 1999), werden mehrere räumlich angeordnete Lautsprecher zur Projektion von Schallquellen verwendet. Mit aufwändigeren Techniken der Klangpräsentation wie z. B. der Wellenfeldsynthese (Horbach und Boone 1999) werden mehrere Lautsprecher so angesteuert, dass eine Klangquelle beliebig im Raum platziert werden kann.

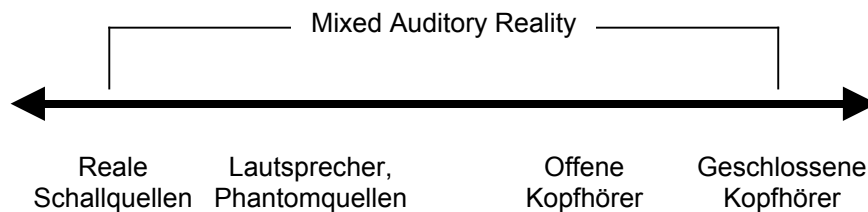


Abb. 31: Kontinuum der Mittel akustischer Darstellungen: von realen Schallquellen (links) bis zur künstlichen Darstellung Kopfhörern (rechts).

Im weiteren Verlauf des Kontinuums dominiert die Verwendung von Kopfhörern und damit auch die größere Trennung zwischen der realen und der virtuellen akustischen Umgebung. Eine weitere Interpretation des Kontinuums in Abb. 31 kann auch die *Unmittelbarkeit der Darstellung* sein, die auf den Merkmalen und Eigenschaften der Präsentationsformen beruht. Reale Klangquellen und Phantomklangquellen können ohne weitere Hilfsmittel von mehreren Zuhörern wahrgenommen werden, während in Richtung der vollständig virtuellen, akustischen Umgebungen die Präsentation individuell erfolgt, in der Regel mit Kopfhörern vermittelt und der Hörer von der realen Hörumgebung getrennt wird.

6.7.2 Architektur der hybriden akustischen Darstellung

Der Ansatz der hybriden akustischen Darstellung besteht darin, die beiden Techniken Lautsprecher- und Kopfhörerpräsentation so zu kombinieren, dass die Eigenschaften beider Präsentationsformen gleichzeitig verwendet werden können. Die hybride akustische Darstellung befindet sich daher in der Mitte des Kontinuums und besteht im Kern aus einem offenen Kopfhörersystem (‘hear-through-display’), dass zusätzlich auch die Wahrnehmung akustischer Signale aus der Umwelt zulässt. Das entspricht einer analogen Anordnung, wie sie für Anwendungen aus dem Bereich der *Augmented Reality* benutzt wird (siehe auch Abschnitt 6.7.1).

Die Gesamtarchitektur der hybriden akustischen Präsentation teilt sich in drei Bereiche (siehe Abb. 33):

- Die physikalische Ebene der Darstellungsmittel,
- Die Ebene der Signalpfadkorrektur und
- Die anwendungsabhängige Vorverarbeitung.

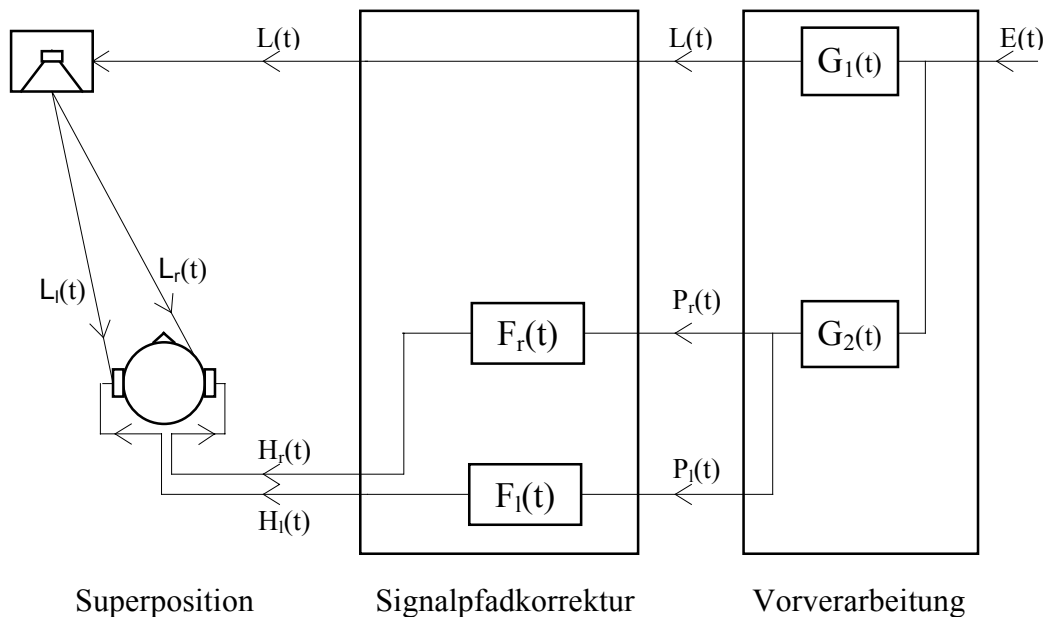


Abb. 33: Gesamtarchitektur der hybride akustische Präsentation: In der Anwendungsebene (rechts) werden z. B. aus einem Eingangssignal drei Zwischensignale $L(t)$, $P_r(t)$ und $P_l(t)$ erzeugt. Nach der Signalpfadkorrektur werden die Kopfhörer- und Lautsprechersignale in der physikalischen Ebene am Ohr des Hörers überlagert.

Der Zuhörer nimmt das Audiosignal $L(t)$, das im Lautsprecher in Schallwellen umgewandelt wird, über die beiden Signalpfade $L_l(t)$ und $L_r(t)$ wahr (der Index l und r verweist hier auf den linken und rechten Anteil). Über das Kopfhörersystem bekommt der Hörer zusätzlich die zwei unabhängigen Signale $H_l(t)$ und $H_r(t)$ zugespielt (siehe Abb. 33). Angenommen wird für die Situation eine ideale Umgebung, in der keine weiteren akustischen Einflüsse wie z. B. durch den Raum eine Rolle spielen. Verzerrungen durch die Schallwandler werden vernachlässigt und zusätzliche Dämpfung durch das offene Kopfhörersystem idealisiert. Des Weiteren liegen die akustischen Signale $L_l(t)$ und $L_r(t)$ nicht explizit als einzelne Signale vor, sondern sind lediglich für die Veranschaulichung und zur Berechnung eingeführt.

Die folgenden Gleichungen beschreiben die zwei Komponenten des resultierenden Signals $S(t)$, die durch die Superposition des Lautsprecher- und Kopfhörersignals an den Ohren des Hörers entstehen:

$$S_l(t) = L_l(t) + H_l(t) \quad (6)$$

$$S_r(t) = L_r(t) + H_r(t) \quad (7)$$

Der Ausgang der weiteren Betrachtungen ist die Identität des Hörerlebnisses einer Schallquelle durch die Signale des Kopfhörers und die des Lautsprechers. Die resultierende Darstellung soll so gestaltet sein, dass die Kopfhörerdarstellung zeitlich und räumlich ‚über‘ der des Lautsprechers projiziert wird. Die dazu notwendige Signalverarbeitung der Kopfhörersignale $H_l(t)$ und $H_r(t)$ besteht aus einer Dämpfung, einer Verzögerung und einer spektralen Anpassung des jeweiligen Signals in Abhängigkeit der Einfallsrichtung des Lautsprechersignals (siehe Abb. 33). Die Entfernung des Lautsprechers zum Kopf des Zuhörers bestimmt die Intensität und die zeitliche Verzögerung der Signale an jedem Ohr.

Die Orientierung des Zuhörers zum Lautsprecher wird im horizontalen und vertikalen Winkel gemessen. Aus diesen Winkeln wird die Art der spektralen Verzerrungen der Signale bestimmt. Auf das Kopfhörersignal $P_l(t)$ und $P_r(t)$ werden die Filter $F_l(t)$ und $F_r(t)$ angewendet, um die Identität des Höreindrucks durch die beiden Signale $H(t)$ und $L(t)$ herzustellen. Diese Verarbeitung des Kopfhörersignals entspricht einer solchen, wie sie im Bereich dreidimensionaler Klangprojektion oft verwendet wird (siehe z. B. Begault 1994; Blauert 1997; Wenzel et al. 2000; Abschnitt 3.1.2, Seite 28). Die Anordnung besitzt für statische Anwendungen Gültigkeit. Sollen darüber hinaus dynamische Konstellationen zwischen Zuhörer und Lautsprecher erlaubt sein, wie sie für interaktive Anwendungen gefordert werden, müssen sich die Parameter der Korrektur mit der Zeit entsprechend ändern. Dazu werden oft sog. Tracking-Systeme verwendet, die die Position und Orientierung des Zuhörers in Echtzeit bestimmen und damit die Korrektur der Kopfhörerdarstellung so steuern können, dass die Illusion einer Klangquelle entsteht, die in einem virtuellen Raum fest positioniert ist (Raumkonstanz, Blauert 1997). In der Ebene der anwendungsabhängigen Vorverarbeitung (siehe Abb. 33 ganz rechts) wird das Eingangssignal $E(t)$ durch die Funktionen $G_1(t)$ und $G_2(t)$ so verarbeitet, dass auf der Zwischenebene zwei Signale $L(t)$, $P_l(t)$ und $P_r(t)$ entstehen, die dann in der physikalischen Ebene übereinander gelagert dargestellt werden. $G_1(t)$ und $G_2(t)$ ergeben sich jeweils aus dem Anwendungskontext der hybriden akustischen Darstellung.

In praktischen Situationen kann die Komplexität der verschiedenen Ebenen der Architektur an die Bedürfnisse der Anwendung angepasst werden. Mögliche Anwendungen der hybriden akustischen Darstellung können in folgenden Bereichen gefunden werden:

- in Kollaborativen virtuellen Umgebungen,
- in Anwendungen aus dem Bereich von Büro- und Fertigungsumgebungen und
- in Anwendungen im Bereich der Unterhaltung, Spiele und Museen.

Soll z. B. eine hybride akustische Darstellung für immersive virtuelle Umgebungen gebildet werden, ist die vollständige und exakte Korrektur der Signale für die Darstellung auf der physikalischen Ebene erforderlich. Üblicherweise existieren in solchen Umgebungen Infrastrukturen für das Erkennen und Verfolgen von Personen im Raum, so dass die Informationen über die Interaktion d. h. über die Position und Orientierung des Benutzers schon vorliegen. In anderen Anwendungssituationen kann es ausreichend sein, dass lediglich die Amplituden und die Zeitverzögerungen der Kopfhörersignale $H_l(t)$ und $H_r(t)$ in der Korrekturebene angepasst werden. Bei der Verwendung von mehreren Lautsprechern, die eine Phantomquelle abbilden sollen, muss berücksichtigt werden, dass damit immer auch ein optimaler Abhörort verbunden ist. Die in Abschnitt 3.1.2, Seite 28, beschriebenen Effekte können den Höreindruck einer festen Position der Phantomquelle stören, wenn man sich aus dem Bereich der optimalen Hörposition bewegt.

6.7.3 Anwendungen für hybride akustische Darstellungen

Als Beispiele für eine Verwendung einer hybriden akustischen Darstellung sollen zwei Situationen beschrieben werden, die die besonderen Eigenschaften einer solchen Darstellung hervorheben:

- die Erweiterung von Objekten der realen Umgebung mit akustischen Signalen und
- die Kooperation in virtuellen akustischen Räumen.

In beiden Fällen spielt es eine besondere Rolle, dass die Darstellung für mehrere Personen gleichzeitig gestaltet ist und die direkte Kommunikation zwischen mehreren Menschen nicht beeinträchtigt wird (interindividuelle Kommunikation). Darüber hinaus steht der Aspekt der individuellen und öffentlichen Darstellung und die Integration realer und virtueller akustische Eigenschaften von Objekten bei den Beispielen im Vordergrund.

Beispiel A:

In einer Gruppensituation, wie sie in Abb. 34 dargestellt ist, ermöglicht die hybride akustische Darstellung einerseits eine ungestörte direkte Kommunikation zwischen den Personen, was durch die Verwendung von offenen Kopfhörersystemen oder auch sog. *in-ear* Systemen begründet wird.

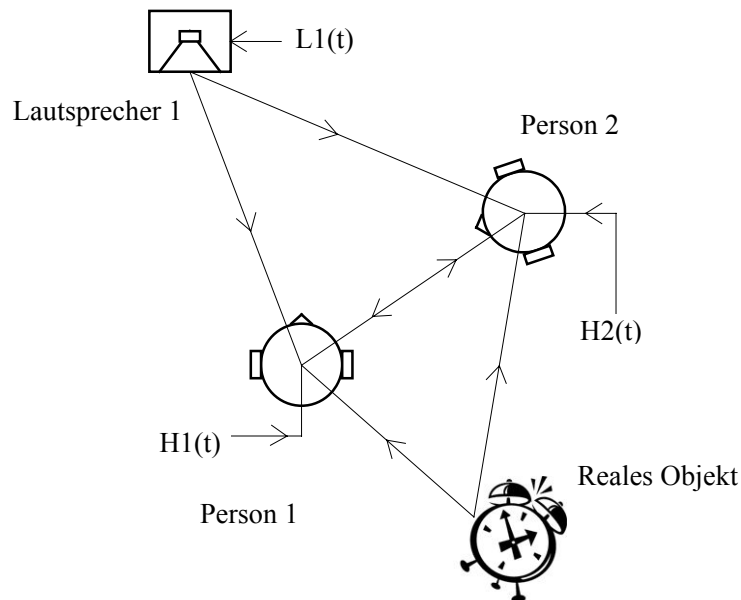


Abb. 34: Die Signalpfade und der bidirektionale Pfad der direkten Kommunikation in einer Anwendung der hybriden akustischen Darstellung. Als reale Schallquellen fungieren ein Lautsprecher und ein Gegenstand, der eigene akustische Signale produziert. Die individuellen Signale $H1(t)$ und $H2(t)$ werden den Personen mit Kopfhörern zugespielt.

Die kleine Bauform der Lautsprecherkapseln der *in-ear* Systeme trägt auch dazu bei, dass die interindividuelle Kommunikation nicht allein schon durch das Tragen eines großen Kopfhörers behindert wird, wie es z. B. in Abowd und Mynatt (2000) für das System *Audio Aura* erwähnt wird. Andererseits erlaubt die hybride akustische Darstellungsform die Kombination von öffentlichen, realen Darstellungen mit persönlichen, virtuellen Klängen. Die folgenden Situationen sollen diesen Sachverhalt illustrieren: In Abb. 34 ist der Wecker Teil der realen Umgebung und die Bewegung des Sekundenzeigers wird durch einen rhythmischen Klang ("tick-tack-tick...") hörbar. Beide Personen nehmen dieses Geräusch als Zeichen (Index) dafür wahr, dass Zeit verstreicht. Der Wecker hat die Funktion, dass er zu einer bestimmten Zeit ein Alarmzeichen gibt, das eine Person z. B. an eine Verabredung für ein Treffen etc. erinnern soll. Solche Alarmsignale können nun in der hybriden akustischen Architektur

individuell über die Kopfhörersignale $H_1(t)$ oder $H_2(t)$ an die jeweilige Person gesendet werden, ohne dass die andere sich im Raum befindende Person den Alarm wahrnimmt. Weiterhin könnte auch ein Lautsprecher im Wecker integriert sein, so dass auch die öffentliche Darstellung des Klangs des Sekundenzeigers kontrolliert werden kann. Mit der Bezeichnung $H_1(t)$ in Abb. 34 werden die beiden Anteile des Kopfhörersignals $H_l(t)$ und $H_r(t)$ zusammengefasst, hier z. B. für die erste Person.

Eine andere Anwendung in diesem Kontext ist z. B. die multilinguale Unterstützung bei einer Filmpräsentation. Die akustischen Signale der Hintergrundkulisse eines Films werden als öffentliche Darstellungen über Lautsprecher präsentiert, während die sprachspezifische Wiedergabe über individuelle Kopfhörer vermittelt wird. Auch die Unterstützung einer erweiterten auditiven Aufmerksamkeit kann mit der hybriden Darstellung geleistet werden. Der Benutzer bzw. Zuhörer wählt eine spezielle Eigenschaft der akustischen Umgebung, an der er besonderes Interesse hat. Individuell werden ihm dann spezielle Geräusche und Klänge vor dem Hintergrund einer allgemeinen akustischen Darstellung hervorgehoben. Durch die implizite Interaktion im Raum, d. h. durch die bloße Zuwendung z. B. zu einem Objekt, kann der Benutzer diese Fokussierung der erweiterten auditiven Aufmerksamkeit steuern.

Beispiel B:

In virtuellen akustischen Umgebungen können nicht nur die Schallquellen virtuell sein, sondern auch der Raum, in dem die Schallquellen platziert sind. Ein virtueller dreidimensionaler Raum mit virtuellen Schallquellen bildet zusammen mit der Möglichkeit der Interaktion des Benutzers bzw. Hörers eine virtuelle akustische Szene. Anwendungen für solche Szenen kommen aus dem Bereich immersiver virtueller Umgebungen, virtueller Raumakustik und musikalisch orientierter Anwendungen (Blauert 1997, Eckel 1999, Jot 1999).

Im Kontext von virtuellen akustischen Räumen erscheint folgende Anwendung für die hybride akustische Darstellung sinnvoll: Wird das Objekt oder der Lautsprecher der Anordnung in Abb. 34 ersetzt durch eine weitere Person, kommt es zu einer Situation, wie sie in Abb. 35 dargestellt ist: Jede Person ist zusätzlich zu ihrem Kopfhörersystem noch mit einem Mikrofon ausgestattet, das entsprechend angebracht, das Sprachsignal der jeweiligen Person aufzeichnet. Spricht dann Person 1 zu Person 2 kann im Weiteren das Signal $M_1(t)$ so verarbeitet werden, dass es für Person 2 den Nachhall eines virtuellen akustischen Raums liefert, in dem sich Person 1 befindet. Unabhängig davon kann aufgrund der symmetrischen Anordnung Person 1 auch Person 2 in einem anderen virtuellen akustischen Raum ‚hören‘. Im Gegensatz zum Beispiel A, in dem die Überlagerung realer und virtueller akustischer Darstellungen an einem Ort im Raum

stattfindet, erfolgt im Beispiel B eine zeitliche Aneinanderreihung von realen und virtuellen Anteilen, d. h. von Direktschallanteil und diffusem Raumnachhall.

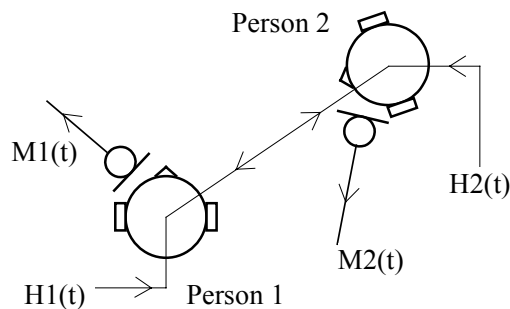


Abb. 35: Die Sprachsignale der Personen 1 und 2 werden z. B. durch drahtlose Mikrofone abgenommen. Diese Signale $M1(t)$ und $M2(t)$ sind die Eingangssignale für eine Verarbeitung, mit der jeweils ein virtueller akustischer Raum gebildet werden kann.

Die Verarbeitung der Mikrofonsignale kann nach üblichen Methoden der Nachhallerzeugung erfolgen (Schröder 1970; Stautner und Puckette 1982; Moorer 1987), wobei insbesondere zu berücksichtigen ist, dass der direkte Signalpfad in der Verarbeitung unberücksichtigt bleiben muss, da er in der hybriden akustischen Darstellung schon als realer, direkter Signalpfad vorhanden ist (siehe Abb. 35).

Eine Betrachtung aus dem Bereich der Raumakustik soll an dieser Stelle verdeutlichen, wie eine einfache Verarbeitung zur Erzeugung von virtuellen akustischen Räumen für die hybride Klangdarstellung erfolgen kann. In Abb. 36 ist eine typische Raumimpulsantwort über der Zeit dargestellt, die oft in drei Komponenten unterteilt wird (siehe z. B. Jot 1999): Direktschall, frühe Reflexionen und später diffuser Nachhall. Mit Direktschall wird der Schallanteil bezeichnet, der auf kürzestem Weg von der Schallquelle zum Zuhörer als Empfänger gelangt. Danach folgen mit geringerer Amplitude die sog. frühen Reflexionen bis diese in den diffusen Nachhall übergehen. Wird ein nachhallfreies Signal z. B. mit einer solchen Raumimpulsantwort gefaltet, werden damit dem Signal die Charakteristiken eines virtuellen Raums aufgeprägt. Andere Verfahren verwenden anstelle der Operation der Faltung Methoden wie z. B. Rückkopplungsnetzwerke.

Der Hauptaspekt für die hybride akustische Darstellung liegt hier in der Aufteilung der Raumimpulsantwort im Zeitbereich, wie es in Abb. 36 mit h_1 und h_2 angedeutet ist. Die Aneinanderreihung der Abschnitte h_1 und h_2 bildet die vollständige Raumimpulsantwort. Für die kooperative Situation, wie in Abb. 35 dargestellt, bedeutet dies, dass Anteil h_1 des Raumeindrucks größtenteils durch die Schallausbreitung im realen Raum gebildet

wird, während Anteil h_2 durch eine Signalverarbeitung des jeweiligen Mikrofonsignals zustande kommen muss. Bei zeitlich korrekten Verhältnissen wird der reale Raumeindruck (Anteil h_1) am Ohr der Hörer ergänzt durch den jeweiligen künstlichen und variablen Nachhall (Anteil h_2). Es kommt zu einer Überlagerung des realen und virtuellen Raumnachhalls. Dabei ist es ausgeschlossen, ein Gesamtsignal $S(t)$ zu erzeugen, das dem Hörer den Eindruck eines kleineren als den realen Raum vermittelt.

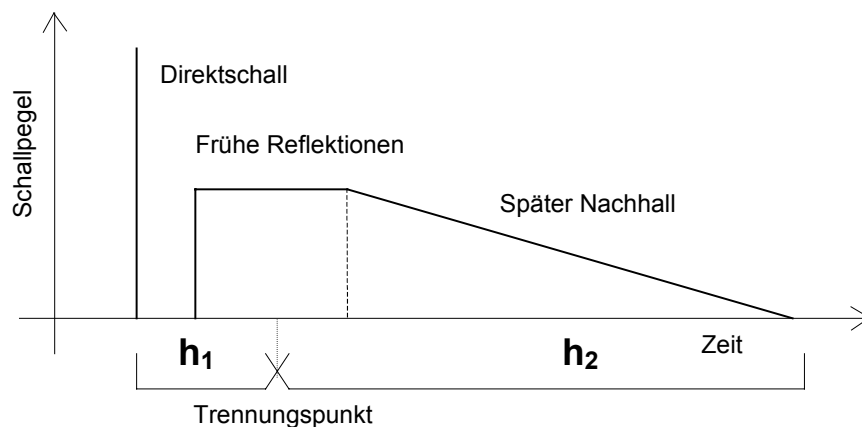


Abb. 36: Schematische Darstellung einer typischen Raumimpulsantwort über der Zeit und eine Aufteilung in zwei Bereiche h_1 und h_2 .

Abschließend soll dieser Ansatz der Aufteilung von Signalanteilen im Zeitbereich einer Aufteilung im Frequenzbereich gegenübergestellt werden. Bei dem Konzept der sog. Subwoofer wird das Signal im Frequenzbereich in zwei Anteile aufgeteilt. Bei einer Anordnung mit zwei Lautsprechern wird der Signalanteil unterhalb von ca. 80Hz von einem zusätzlichen dritten Lautsprecher abgestrahlt. Der Ort dieses Subwoofers kann von Menschen aufgrund der großen Wellenlängen in diesem Bereich nicht eindeutig lokalisiert werden und weshalb ein Schallwandler für diesen Frequenzbereich im Raum verwendet wird. Beim Ansatz der Aufteilung von Signalanteilen im Zeitbereich wie bei der hybriden Präsentation existiert ebenso eine räumliche Aufteilung der Lautsprecher, weshalb hier von einem *firstspeaker* und *latespeaker* gesprochen werden soll. Der *firstspeaker* wird durch den Lautsprecher gebildet, während der Kopfhörer den *latespeaker* darstellt. Die Aneinanderreihung der Schallanteile beider *speaker* erfolgt dann an den Ohren des Hörers. Die Gemeinsamkeiten des Subwoofer-Ansatzes und der hybriden Präsentation sind die Aufteilung in mehrere Signalanteile, die räumlich verteilte Beschallung und die Wiederzusammenführung der einzelnen Schallanteile, so dass es zu einem plausiblen Schallereignis beim Benutzer führt.

7 Entwicklungen zur Echtzeit-Signalverarbeitung auf Standardcomputern

Die Integration akustischer Repräsentationen in Computeranwendungen erfolgt in ähnlicher Weise wie die Integration optischer Darstellungen. Eine Programmierschnittstelle akustischer Darstellungen dient dem Entwickler von Anwendungen dazu die akustische Darstellung benutzen zu können, während die Erzeugung der Darstellung unterhalb dieser Schnittstelle erfolgt. Im Folgenden wird eine Programmierbibliothek beschrieben, mit deren Hilfe es möglich ist, die Erzeugung der akustischen Signale für die Darstellung in eine Anwendung zu integrieren. Die Bibliothek stellt elementare Objekte der akustischen Signalerzeugung zur Verfügung, die zur Verwendung aus der Anwendung heraus zu einer Einheit bzw. zu einem Objekt zusammengestellt werden. Es wird dadurch möglich, kurze Zeitverzögerungen zwischen einem Eingabeereignis in der Anwendung und der Erzeugung eines akustischen Signals zu realisieren (Gestaltungsregel D, Seite 62). Die akustische Darstellung bei einer Interaktion z. B. mit einem Stift auf einem Grafiktablett erscheint dem Benutzer dann nahezu synchron mit seinen realen Handbewegungen (siehe Abschnitt 6.4.4). Ereignisse, die aus Audiosignalen heraus durch Analyseverfahren im Frequenz- und Zeitbereich extrahiert werden, können ebenso mit einer geringen Verzögerung an die Anwendung weitergereicht werden (z. B. Tonhöhendetektion). So ist die zügige Weiterverarbeitung von abgeleiteten Ereignissen in Anwendungen gewährleistet. In einem weiteren Teil dieses Kapitels wird die Leistungsstärke von Standardcomputern für Algorithmen der Signalverarbeitung untersucht und der besondere Fall der Realisierung einer Signalverarbeitung in parallelen Prozessen näher beschrieben. Der verwendete Algorithmus ermöglicht die effiziente Faltungsoperation im Frequenzbereich und bietet gleichzeitig günstige Eigenschaften für den Einsatz in interaktiven Echtzeitanwendungen. Eine hardwareunabhängige Realisierung auf Standardcomputern wird diskutiert.

7.1 Bibliothek von Objekten zur Signalverarbeitung

Damit eine einfache und effiziente Anbindung von akustischen Darstellungen in interaktive Anwendungen erfolgen kann, wird eine Programmierbibliothek von Objekten zur Signalgeneration und -verarbeitung entwickelt. Die Eigenschaften bisheriger Ansätze reichen teilweise nicht aus, um die Anforderungen einer geringen Ausgabeverzögerung und der effizienten Erzeugung zu erfüllen. Die Programmierbibliothek zur Behandlung von Audiosignalen in Computerumgebungen in Echtzeit ist in Objektklassen aufgebaut (Real-Time Signal Object, *RTSigObj*), die

ähnliche Eigenschaften wie die Klassen objektorientierter Programmierung besitzen. Ein Vorteil dieser Art der Programmierung ist die systematische Erstellung von Objekten in einheitlicher Weise bei gleichzeitiger effizienter Verarbeitung der Signale durch die Objekte. Auf die Verwendung der objektorientierten Eigenschaften von Computersprachen wie z. B. C++ oder Objective-C wurde verzichtet. Ausschließlich die relevanten Eigenschaften einer solchen Programmiersprache wurden in ANSI-C nachgebildet. Dadurch bleibt die Ausführung einer ‚Methode‘ dieser Objekte immer eine Ausführung einer direkten Funktion, ohne dass zusätzlich Methodentabellen zur Auswertung der Funktionsadresse während der Laufzeit ausgewertet werden müssen. Da z. B. bei einer Blockverarbeitung mit der Blocklänge von 64 Abtastwerten bei einer Abtastrate von 44.1 kHz der Aufruf der Verarbeitungsfunktion alle 1.45 ms erfolgt, kann durch den Funktionsaufruf gegebenenfalls Verarbeitungsleistung eingespart werden. Dieser Vorteil der direkten Funktionsaufrufe beträgt bei älteren Prozessortypen (Pentium II oder älter) bis zu 6% gegenüber den Methodenaufrufen in C++ vorhanden. Bei neueren Prozessortypen hingegen ist dieser Vorteil nicht mehr deutlich erkennbar. Denkbar ist auch eine gemischte Form der Programmierung von Objekten, so dass die Klassenbeschreibung in objektorientiertem C++ erfolgt, die Funktionen aber, die in Echtzeit benutzt werden, als C-Funktionen ausgeführt werden.

Im Gegensatz zu Ansätzen wie dem Synthesis-ToolKit (STK, Cook und Scavone 1999) erlauben die Objekte der *RTSigObj* Bibliothek sowohl die Verarbeitung eines Blocks von Abtastwerten bei einem Aufruf zur effizienteren Verarbeitung als auch den Aufruf der Verarbeitung pro Abtastwert. Mit diesem Aufruf sind dann ‚Verschaltungen‘ von Signalobjekten mit minimaler Verzögerung in den Zweigen von Signalarückkopplungen möglich (rekursive Strukturen). Eine derartige Programmierung ist z. B. bei blockverarbeitenden Systemen wie *Max* (Max/MSP 1997) oder *PD* (Puckette 1996) auf der Anwendungsebene nicht möglich. Oft müssen bei diesen Systemen sog. externe Objekte programmiert werden, um die vollständige Kontrolle über die Verarbeitung der Abtastwerte zu erlangen. Häufig wird dieses Problem bei der Realisierung von Kammfiltern bzw. Verzögerungsketten deutlich. Für den hier verwendeten Ansatz wird die Ein- und Ausgabe der Audiodaten die Programmierbibliothek *PortAudio* verwendet (Bencina und Burk 2001), die für verschiedene Betriebssysteme implementiert ist. Abhängig von der jeweiligen Hardware werden jedoch sehr unterschiedliche Ein- und Ausgabeverzögerungen erzielt (MacMillan et al. 2001). Auf eine Kommunikationsschnittstelle zur Kontrolle aus anderen Programmen mithilfe einer Socket-Programmierung wie z. B. bei *STK* wird verzichtet, da es Ziel der Bibliothek ist, die Signalverarbeitung und -erzeugung in Anwendungen direkt zu integrieren. Der Polymorphismus objektorientierter Programmiersprachen wird für die Entwicklung der

RTSigObj Bibliothek nachgeahmt, so dass z. B. alle Objekte über eine *new*, *init* und *free* Methode bzw. Funktion verfügen und die Funktionen *block* und *sample* für die Signalverarbeitung in Echtzeit verwendet werden. Auf die Nachbildung der Fähigkeit der Vererbung von Klassen wurde verzichtet, da bis auf wenige Ausnahmen hierarchische Strukturen in den Objekten nicht zu erwarten sind, bei denen sich eine solche Eigenschaft objektorientierter Programmierung auszahlen würde. Die Darstellung der einzelnen Abtastwerte in der Bibliothek erfolgt in einer 32-Bit Fließkomma-Darstellung und bedeutet einen Kompromiss zwischen Speicherplatz, Auflösung und Verarbeitungsgeschwindigkeit der Abtastwerte (Puckette 1991a, Freed 1993, Bencina und Burk 2001).

7.1.1 Integration in Anwendungen

Die Integration einer Signalgeneration und -manipulation für akustische Darstellungen in Computeranwendungen erfolgt über die Einbindung von Funktionsbibliotheken. Die Programmierbibliotheken *RTSigObj* und *PortAudio* (2001) bilden die Grundlage zur Verarbeitung, Kontrolle und Manipulation der Audiosignale in Echtzeit. Beide Bibliotheken sind in der Programmiersprache ANSI-C geschrieben. Die zentrale Funktion, in der die Signalverarbeitung erfolgt, wird durch einen Funktionszeiger an die Eröffnungsroutine der *PortAudio* Bibliothek übergeben und dort in einem Unterprozess periodisch ausgeführt. Es können Objekte der Klangerzeugung (Modelle) gebildet werden, in denen die Erzeugung und Verschaltung der Signalobjekte aus der Bibliothek *RTSigObj* erfolgt. Die Steuerung der Objekte der Signalverarbeitung erfolgt während der Laufzeit direkt über die Felder der Objekte der *RTSigObj* Bibliothek. Dazu kann beispielsweise auf der Applikationsebene ein Schieberegler der grafischen Benutzungsoberfläche die Ausgangslautstärke eines Signals manipulieren. Die Auswirkungen eines solchen Ereignisses werden nur im Takt der Blockverarbeitung am Ausgang der Audioverarbeitung hörbar. Dies kann dann zu Problemen führen, wenn die Blocklänge zu groß ist und die Auswirkungen auf das Audiosignal im ungünstigsten Fall, bis um die Zeitdauer eines Blocks verzögert werden (siehe auch Abschnitt 4.2.1, Seite 43). Zusammenhängende optisch-akustische Darstellungen würden dadurch zusätzlich in der Möglichkeit ihrer kohärenten Wahrnehmung und Wirkung beeinträchtigt. In solchen Situationen wäre zu überlegen, ob das Ereignis nicht besser als Signalstrom in die Verarbeitung eingespeist wird, was jedoch zusätzlichen Aufwand und Verarbeitungszeit in Anspruch nimmt. Üblicherweise wird die Blocklänge auf 64 Abtastwerte gesetzt (Puckette 1991a). Dadurch ergibt sich für die zeitliche Auflösung der Ereignisverarbeitung eine Zeitdauer von $T_e = 64/f_s$. Bei einer Abtastrate von 44.1 kHz erfolgt die Ereignisverarbeitung mit der Rate $f_e = 689$ Hz. Im ungünstigsten Fall

würde dann ein Versatz von 1.45 ms zwischen Ereignis und Auswirkung im Audiosignal die Folge sein. Diese Zeitspanne liegt noch unterhalb der Auflösungsschwelle der auditiven Wahrnehmung von Ereignissen und kann daher noch toleriert werden (siehe Abschnitt 3.1.1, Seite 26).

Um für die Signalverarbeitungsfunktion die Objekte miteinander zu ‚verschalten‘, wird Speicherplatz benötigt, der Signalblöcken zugewiesen wird. Die Signalblöcke dienen der Verbindung des Ausgangs eines Signalobjekts mit dem Eingang des nächsten Objekts. Ähnlich wie in einem gerichteten Graphen sind die gerichteten Kanten Signalverbindungen, während die Knoten Signalobjekte repräsentieren (Puckette 1991b). Einfache arithmetische Operationen wie Addition und Multiplikation können auf zwei Signalblöcke direkt angewendet werden, ohne dass dazu ein weiteres Objekt benötigt wird.

7.1.2 Objektgruppen

Die Objekte, die für die *RTSigObj* Bibliothek implementiert wurden, bestehen aus elementaren Objekten zur Erzeugung und Manipulation von Signalen, zusammengesetzten Objekten und Objekten zur Filterung von Audiosignalströmen. Alle Objekte besitzen einen Verweis auf ein Ursprungsobjekt (*RTSigObj*), das keine Verarbeitungsfunktion, dafür jedoch Felder besitzt, über die jedes weitere Objekt die aktuelle Abtastrate und die Blocklänge der Verarbeitung auslesen kann. Darüber hinaus stellt dieses Ursprungsobjekt eine Zeigervariable zur Verfügung, über die alle Objekte der Bibliothek Informationen in eine Datei oder auf den Konsolenbildschirm ausgeben können. Das *RTSBlock* Objekt stellt Speicherplatz für einen Signalblock in der aktuellen Blocklänge zur Verfügung, mit dem die Übertragung der Signale zwischen dem Ausgang eines Objektes und dem Eingang eines folgenden Objektes bewerkstelligt wird.

Die folgende Beschreibung der Objektgruppen gibt den Stand der gebildeten Klassen wieder. Es handelt sich neben den Wurzelklassen um die Gruppen: Generator, Filter, Effekt und Modell. Mit jeder neuen Anwendung können schnell benötigte Objekte entwickelt und in die Bibliothek aufgenommen werden, so dass sie für nachfolgende Entwicklungen und Anwendungen zur Verfügung stehen. Zusammengesetzte Objekte, die aus mehreren verschalteten Objekten bestehen, können nach dem Abschluss ihrer Entwicklung ebenso in die Bibliothek mit aufgenommen werden.

Gruppe	Klasse	Beschreibung
Basisklasse	RTSigObj	Objekt für allgemeine Eigenschaften der Verarbeitung
	RTSBlock	Signalblockobjekt zur Verknüpfung von Objekten

Tabelle 1: Beschreibung der Basisklassen

Die Gruppe der Generatoren enthält Objekte, die Signale erzeugen, welche durch Kontrollereignisse in ihrer Charakteristik verändert werden können. So z. B. kann das RTSSine Objekt, das als Ausgangssignal eine Sinusschwingung erzeugt, in der Frequenz manipuliert werden. Mit dem RTSNoise Objekt steht ein Rauschgenerator zur Verfügung, der aus einer Zufallsvariablen des Computersystems ein gleichverteiltes oder gaußverteiltes weißes Rauschen annähert. Weiterhin besteht die Möglichkeit auch sog. rosa Rauschen (pink noise) mit dem Objekt zu erzeugen. Die Gruppe der Generatoren wird durch Objekte zur Erzeugung von Rechteckschwingungen und zum Einlesen von Audiodateien ergänzt.

Gruppe	Klasse	Beschreibung
Generator	RTSSine	Sinusgenerator nach der <i>Table-LookUp Methode</i>
	RTSNoise	Allgemeiner Rauschgenerator für weißes und rosa Rauschen
	RTSSndFile	Objekt zum Einlesen und Schreiben von Audiodateien
	RTSPulse	Parametrischer Rechteckgenerator
	RTSGaussPulse	Komplexer Pulsgenerator, bei dem Amplitude und Pulsweite stochastischen Einflüssen unterliegen

Tabelle 2: Beschreibung der Generatorklassen

Alle Filterobjekte basieren auf dem RTSBiquad Objekt, das eine generelle Filterstruktur zweiter Ordnung mit zwei Polen und zwei Nullstellen ermöglicht. Alle anderen Filterobjekte wie RTSLpass1, RTSShelv usw. basieren auf diesem Filterobjekt und zeichnen sich durch zusätzliche Funktionen aus, mit denen die

Koeffizienten für das spezielle Filterdesign berechnet werden. Mit dem sog. Shelvingfilter-Objekt `RTSShelv` steht z. B. in der *RTSigObj* Bibliothek ein oft verwendetes Filter zur Audioverarbeitung zur Verfügung, mit dem leicht klangliche Veränderungen realisiert werden können (Zölzer 1997).

Gruppe	Klasse	Beschreibung
Filter	<code>RTSBiquad</code>	Zentrales digitales Filter mit 3 Zähler- und 3 Nennerkoeffizienten, Filter zweiter Ordnung
	<code>RTSLpass1</code>	Parametrisches Tiefpassfilter erster Ordnung, basierend auf <code>RTSBiquad</code>
	<code>RTSLpass2</code>	Parametrisches Tiefpassfilter zweiter Ordnung
	<code>RTSShelv</code>	Parametrisches Shelvingfilter zweiter Ordnung für Tief-, Hoch- und Bandpassfilterungen

Tabelle 3: Übersicht über die einfachen Objekte zur Filterung

Gruppe	Klasse	Beschreibung
Effekt	<code>RTSDelay</code>	Objekt zur variablen Zeitverzögerung von Signalströmen
	<code>RTSComb</code>	Parametrische Kammfilterstruktur mit vorwärts und rückwärts gerichteter Verzögerung
	<code>RTSDecorr</code>	aus fünf Allpassfiltern zur Dekorrelation von Signalen bestehendes Objekt
	<code>RTSVarRate</code>	Objekt zur Veränderung der Abtastraten mit linearer Interpolation

Tabelle 4: Übersicht über die Effektoobjekte

Die Objekte zur Erzeugung von besonders einfachen Effekten basieren hauptsächlich auf der Verzögerung von Signalwerten und deren Rückkopplung. Dazu wird im Objekt Speicherplatz zugewiesen, der als ein Ringspeicher die eingehenden Signalblöcke aufnimmt. Die maximale Verzögerung ist durch die Länge des zirkularen Signalspeichers begrenzt und beim Überlauf werden die gespeicherten Signalblöcke überschrieben. Hier kommt es häufig zur Unterschätzung des Speicherbedarfs solcher zirkularen Signalspeicher. Soll es möglich sein z. B. ein Audiosignal mit 16 Bit Auflösung und einer Abtastrate von 44.1 kHz bis zu 1 sec. zu verzögern, so muss Datenspeicherplatz der Größe 86.1 kByte bereitgestellt werden. Für die meisten Anwendungen auf Standardcomputersystemen sind diese Anforderungen jedoch leicht zu erfüllen, lediglich die Geschwindigkeit des Speicherzugriffs kann zu Problemen führen.

Die Modellgruppe stellt eine besondere Gattung von Objekten dar. Die in einem solchen Objekt realisierte Signalgeneration bzw. -manipulation ist zusammengesetzt aus mehreren Objekten, die elementare Erzeugung und Verarbeitung leisten. Man kann das Modell auch als einen komplexen Generator bezeichnen, der mit verschiedenen Parametern gesteuert wird. Ähnlich wie bei den ‚physikalischen‘ Modellen (Cook 1997) wird z. B. durch das `RTSParticleModel` mit einfachen und effizienten Methoden die Imitation der akustischen Eigenschaften bei der Interaktion auf Oberflächen ermöglicht. Wie in Abschnitt 6.4.4 beschrieben sollen diese Modelle die Invarianzen von Oberflächen plausibel vermitteln. Das `RTSPenModel` hingegen ist aus dem Resonanzmodell eines Schreibstiftes abgeleitet und verhält sich als passiver Resonator eher wie ein Frequenzfilter (Dutilleux und Müller-Tomfelde 1999).

Gruppe	Klasse	Beschreibung
Modell	<code>RTSParticleModel</code>	Komplexes Modell zur Erzeugung von stochastischen Geräuschen von Partikeln
	<code>RTSPenModel</code>	Modell des Resonanzverhaltens eines Schreibstiftes
	<code>RTSTube</code>	Kammfilter mit Tiefpassfilter zur Imitation von Rohrresonanzen

Tabelle 5: Übersicht über die Modellobjekte

7.1.3 Makro-Umgebung zur standardisierten Programmierung

Beim Entwurf der *RTSigObj* Bibliothek wurde versucht, die Ansätze und Vorteile von objektorientierter Programmierung (OOP) auszuschöpfen, auch ohne eine Programmiersprache wie C++ oder Objective-C zu verwenden. Dafür wurden Makro-Funktionen entwickelt, die es ermöglichen, jedes Objekt der Bibliothek ähnlich wie in OOP Sprachen in einer standardisierten Form zu beschreiben. Die vereinfachte OOP (portierbare oder auch pseudo-OOP, abgekürzt *poop*) entstand mit dem Ziel, einerseits eine effiziente und portierbare Entwicklung zu realisieren und dabei andererseits nicht auf die Möglichkeiten und Vorteile einer strukturierten Beschreibung und Entwicklung von Klassen zu verzichten. Der Zugriff auf die Datenfelder ist weniger einfach gelöst als bei Sprachen der OOP. Jedoch ist der Zugriff auf die Felder über einen Zeiger auf die Klasseninstanz einheitlich und in allen Methoden gleich. Die Definition der Methoden ist ebenfalls angelehnt an die Art, wie sie in OOP Sprachen erfolgt. Verzichtet wird jedoch auf die Anlage von Methodentabellen, so dass ein Methodenaufruf eines *poop*-Objektes einem C-Funktionsaufruf gleichkommt, wobei der erste Parameter der Funktion immer ein Zeiger auf die Datenstruktur der Zielinstanz ist.

7.1.3.1 Klassendefinition

Die sog. *poop*-Headerdatei einer solchen Klasse ist in drei Bereiche aufgeteilt: die Definition des Namens, die Definition der Datenfelder bzw. Datenstruktur und die Definition der Methoden. Mit dem Objekt *RTSLpass1* soll die Verwendung veranschaulicht werden:

```
1  #include "poopinit.h"
2      #define CLASS_NAME RTSLpass1
3  #include "poopclass.h"
4
5  CLASS_PREAMBLE
6
7  DATA_DEF
8      RTSigObj *rtso;
9      RTSBiquad *rtsb;
10     float F,K;
11  DATA_END
```

```
12
13  /* standard methods */
14  extern CLASS_NAME *METHOD(new) (void);
15  extern int METHOD(free) (IARG);
16  extern int METHOD(init) (IARG);
17
18  /* control methods */
19  extern int METHOD(setF) (IARG, float f);
20  extern int METHOD(calc) (IARG);
21
22  /* signal processing methods */
23  extern int METHOD(block) (IARG, float *in, float *out, int n);
24  extern int METHOD(sample) (IARG, float *in, float *out);
```

Im ersten Bereich (Zeile 1-3) wird der Makro-Variable `CLASS_NAME` für die Zeit der Übersetzung dieser Programmdatei der Name der Klasse zugeordnet, auf die sich alle weiteren Makrofunktionen beziehen. Dies geschieht durch das Einbinden der Makro-Dateien `poopinit.h` und `poopclass.h` und die Klammerung der Makro-Definition `CLASS_NAME`. Der Name der Klasse wird dadurch als globale Variable während der Vorübersetzung der Klasse für alle Makros bereitgestellt, so dass z. B. Funktionsnamen oder Klassenstrukturen ein eindeutiges Klassensymbol erhalten, bevor die Datei kompiliert wird. Es folgt die Definition der Klassenfelder, die mit dem Makro `DATA_DEF` (ab Zeile 7) eingeleitet, wird und endet mit dem Schlüsselwort `DATA_END`. In diesem Bereich können sämtliche ANSI-C kompatiblen Definitionen als Felder des Objekts angeführt werden. Im dritten Bereich der Header-Datei werden die Methoden deklariert. Standardmethoden wie `new`, `free` und `init` sind für alle Klassen vorhanden sowie die Methoden `block` und `sample` zur Verarbeitung von Signalen in Echtzeit.

Nach der Vorkompilierung hat die Datenstruktur der Klasse folgende Form:

```
1  typedef struct
2  {
3      char *_isa;
```

```
4      RTSigObj *rtso;
5      RTSBiquad *rtsb;
6      float F,K;
7  } RTSLpass1;
```

Der neue Datentyp mit dem Namen `RTSLpass1` ist eine einfache Struktur, die die Felder der Klasse zusammenfasst und für die weitere Programmierung zur Verfügung stellt. Zusätzlich fügt das Makro `DATA_DEF` der Datenstruktur noch eine Variable (`_isa`) hinzu, mit zur Laufzeit dann Informationen über die Klasse als Zeichenfolge zur Verfügung stehen. Bei der Erzeugung einer neuen Instanz mit der Klassenmethode `new` wird immer der Zeiger auf eine dynamisch angelegte Variable zurückgegeben.

7.1.3.2 Implementierung der Methoden

Die Header-Datei dient als Schnittstelle für die Verwendung des Objektes in der Bibliothek und aus Anwendungsprogrammen heraus. Sie wird von der Programmdatei eingebunden, die die Methodenimplementierung enthält. Alle Methoden bzw. Funktionen der *poop*-Klassen haben folgende Charakteristika:

- Das Symbol der Funktion wird bei der Übersetzung zusammengesetzt aus dem *poop*-Methodenname wie z. B. `new` oder `init` und dem Klassennamen. So ist dann die *poop*-Methode `init` eines Objektes der Klasse `RTSLpass1` nach der Kompilierung mit dem Symbol `initRTSLpass1` für andere Programmbereiche verfügbar.
- Die erste Variable, die den Funktionen beim Aufruf übergeben werden muss, ist ein Zeiger auf die zu dem Objekt gehörende Datenstruktur und wird mit dem Makro `IARG` (Instanzargument) abgekürzt.

Als Beispiel der Implementierung sollen hier zwei Methoden dargestellt werden, um die Formen des Zugriffs auf die Datenfelder des Objektes und den Aufruf der Methoden innerhalb eines Objektes zu demonstrieren. Mit der ersten Methode `setF` wird die Grenzfrequenz des Filterobjektes festgesetzt und danach wird die Methode `calc` zur Berechnung der Filterkoeffizienten aufgerufen:

```
1  int METHOD(setF)(IARG, float f)
2  {
3      ivar(F) = f;
4      METHOD(calc)(self);
```

```

5         return 0;
6     }
7
8     int METHOD(calc) (IARG)
9     {
10         ivar(K) = (float) (ivar(F)/ivar(rtso)->sr*2.0);
11         ivar(rtsb)->a[0] = ivar(K);
12         ivar(rtsb)->b[1] = (float) (ivar(K) - 1.0);
13         return 0;
14     }

```

Mit der Makro-Funktion `ivar()` wird auf die Variable der Instanz verwiesen (z. B. Zeile 3). Der Aufruf der Methode `calc` in Zeile 4 erfolgt mit dem Argument `self`. Diese Variable ist in jeder Methode vorhanden und erlaubt den Zugriff auf die Datenstruktur der aktuellen Klasseninstanz. Diese Variable entspricht der `this` Variablen in C++ bzw. der `self` Variablen in Objective-C. Über das Wurzelobjekt im Datenfeld des Objektes `rtso` wird in Zeile 10 auf die Abtastrate (sampling rate, `sr`) des Gesamtsystems zugegriffen.

Nach der Vorkompilierung mit den *poop*-Makrodefinitionen ist die Methode der *poop*-Klassen eine C-Funktion, die wie im Fall der Methode `setF` wie folgt aussieht:

```

1     int setFRTSLpass1(RTSLpass1 *self, float f)
2     {
3         self->F = f;
4         calcRTSLpass1(self);
5         return 0;
6     }

```

Zu bemerken ist an dieser Stelle, dass bei der Implementierung der Methode mit den *poop*-Makrodefinitionen der Klassenname wie in diesem Fall `RTSLpass1` nicht explizit angegeben wird bzw. erscheint. Der Name der Klasse wird ausschließlich einmal in der Headerdatei angegeben und ist dann für die Übersetzung des Objektes

gültig. Mit einem sog. Template ist eine grundlegende Struktur eines *RTSig*-Objektes vorgegeben und für eine neue Klasse wird dann lediglich der Name der Klasse geändert und an entsprechenden Stellen der spezielle Programmcode in die Funktionen eingesetzt.

7.2 Beispiele zur Signalgeneration und -manipulation

Um die Verwendung der *RTSigObj* Bibliothek zu demonstrieren, sollen hier zwei Arten von Modellgruppen beschrieben werden, die für die Beispiele der akustischen Darstellungen in Kapitel 6 verwendet wurden. Die Entwicklung der Modelle ist stellenweise durch physikalische Größen und Gesetze vorgegeben. Eine physikalische Analyse der Ursachen von akustischen Prozessen dient immer als Grundlage bzw. Ausgangspunkt für die Gestaltung von Generatoren, die den Klang imitieren sollen (Gaver 1993a). Auch bei der Schaffung von abstrakten Geräuschgeneratoren ist eine Anlehnung an physikalische Objekte mit bekannten akustischen Eigenschaften als Basis zur Entwicklung sinnvoll, um plausible Klangeffekte zu erzeugen.

7.2.1 Physikalische Modelle

Sollen z. B. die akustischen Eigenschaften von realen geometrischen Objekten imitiert werden, dann können durch die Abmaße der Objekte Aussagen z. B. über das Resonanzverhalten gemacht werden. Im Gegensatz zu geschlossenen mathematischen Lösungen der Wellenausbreitung über der Zeit steht bei diesen Modellen lediglich die Grundcharakteristik wie z. B. die Resonanzfrequenz im Vordergrund (Dutilleux und Müller-Tomfelde 1999). Im Sinne der *Ecological Acoustics* (siehe Abschnitt 3.2, Seite 32) sollen die Modelle die hörbaren Anteile erzeugen, die der Zuhörer als eine Grundeigenschaft des Objektes wahrnimmt. Das Modell soll durch diese Charakteristik dem Hörer plausibel erscheinen und auf eine exakte Simulation auch zu Gunsten einer effizienten Erzeugung sollverzichtet werden.

Eine zentrale physikalische Eigenschaft für die Modellierung von schallenden geometrischen Objekten sind sog. stehende Wellen. Solche Wellen sind Schallwellen, die zwischen zwei räumlichen Begrenzungen reflektiert werden. Alle harmonischen Obertöne dieser Grundmode passen ebenfalls in die Begrenzung und es entsteht ein streng harmonischer Resonator. Typische Beispiele solcher Resonatoren sind z. B. Rohre, aber auch parallel verlaufende Wände von Räumen (siehe Abschnitt 6.4.2). In der Regel werden die akustischen Charakteristiken von Resonatoren durch Verzögerungsketten modelliert, bei denen ein Teil der Energie des Eingangssignals wieder zurück in das System reflektiert wird, während der andere Teil aus dem System

ausgekoppelt wird. Bei der Modellierung eines Rohrs entspricht die Länge des Rohrs der Resonanzfrequenz eines sog. Kammfilters. Zusätzlich kann in die Rückkopplung des Systems noch ein Tiefpassfilter eingesetzt werden, mit dem der Einfluss des Durchmessers des Rohrs modelliert wird (Dutilleux und Müller-Tomfelde 1999). Eine derartige Struktur zur Modellierung des Resonanzverhaltens eines Rohrs wurde mit der Klasse `RTSTube` in der Bibliothek *RTSigObj* realisiert.

Die rekursive Kammfilterstruktur eignet sich besonders gut für eine effiziente Modellierung, da mit einfachen Verarbeitungsoperationen komplexe Klänge und Effekte hergestellt werden können (z. B. Erzeugung von Nachhall mit *Feedback Delay Networks*, Jot 1992). Sollen auch Verzögerungswerte zwischen zwei diskreten Werten einstellbar sein, so muss zwischen den zeitdiskreten Signalwerten interpoliert werden. Der Aufwand der Interpolation addiert sich zu den Verarbeitungsoperationen hinzu und trägt im Fall z. B. einer quadratischen Interpolation bedeutend zum Aufwand der Berechnung bei (Rocchesso 1998).

7.2.2 Abstrakte Modelle

Neben den physikalischen Modellen, bei denen es einen direkten Zusammenhang zwischen einer Größe eines realen Vorbildes und einem Parameter im Modell gibt, soll eine weitere Gruppe von Modellen beschrieben werden. Die Modelle sollen als abstrakt bezeichnet werden, wenn kein unmittelbarer Zusammenhang existiert, der auf einer messbaren physikalischen Größe beruht. Als Ergebnis einer akustischen Analyse, die mit den Ohren oder messtechnischen Hilfsmitteln erfolgt, soll mit dem abstrakten Modell ein einfaches und effizientes Syntheseverfahren geschaffen werden. Ziel des abstrakten Modells ist es wie beim physikalischen Modell, charakteristische akustische Merkmale des realen Geräuschs herauszuarbeiten und zu imitieren. Entscheidend bei den abstrakten Modellen ist, dass ein gewisser Anteil an Kreativität bei der Entwicklung nötig ist, um ein geeignetes akustisches Modell zu schaffen. Anhand des Beispiels der akustischen Darstellung für die Interaktion mit einem Stift auf einer Oberfläche soll der Generator beschrieben werden, der dem Benutzer verschiedene Oberflächenbeschaffenheiten über die akustische Darstellung vermittelt.

Der zentrale Parameter dieser Interaktion ist die Geschwindigkeit, mit der der Stift über die Oberfläche bewegt wird. Wichtig ist, dass der vermittelte Eindruck einer Oberflächenqualität bei der Interaktion konstant bleibt, während die Erzeugung bei der Stiftinteraktion dynamisch ist. Es handelt sich um eine sog. strukturelle Invarianz des Objekts (siehe Abschnitt 3.2.1, Seite 34), die der Benutzer aktiv erfahren kann.

Kern der Synthese ist ein parametrischer Rechteckgenerator, der in seiner Frequenz und Pulsweite kontrolliert werden kann. Zusätzlich kann die Amplitude des Rechtecksignals eingestellt werden. Die Frequenz wird proportional durch die Stiftgeschwindigkeit gesteuert und vermittelt dadurch eine Invarianz der Struktur der Oberfläche (Meso-Textur, siehe Abschnitt 6.4.4, Seite 81). Wird der Stift langsam bewegt, finden wenige Kollisionen mit den Unebenheiten auf der Oberfläche statt und es ertönen wenige Impulse des Generators, während schnelle Bewegungen viele Impulse erzeugen. Dadurch wird das mentale Modell einer gleichmäßig strukturierten Oberfläche beim Benutzer bewirkt.

Über zwei sog. Gauß-Prozesse werden die Parameter Pulsweite und Amplitude des Rechteckgenerators um stochastische Eigenschaften ergänzt. Damit wird es möglich, unterschiedlich große Unregelmäßigkeiten der virtuellen Oberflächenqualität zu imitieren. In einer bildlichen Analogie kann davon gesprochen werden, dass die Amplitude des Generators mit der Höhe und die Pulsweite mit der Breite von Partikeln auf der Oberfläche korrespondieren. Das Ausgangssignal wird zusätzlich durch einen Tiefpass gefiltert, so dass das Auf- und Absteigen der Flanken des Rechtecksignals weniger abrupt erfolgt (siehe Abb. 25, Seite 83). In der *RTSigObj* Bibliothek bildet diese Art von Generator die Grundlage der Klasse `RTSParticleModel`.

7.3 Leistungsverhalten von Standardcomputersystemen

Die Bewertung der Leistungsfähigkeit von Standardcomputern bei der Bearbeitung von Aufgaben der Signalverarbeitung soll an Hand des Algorithmus zur schnellen Fourier-Transformation (fast Fourier transformation, FFT) untersucht werden (Kammeyer und Kroschel 1989). Dieser Algorithmus hat verschiedene Eigenschaften, die ihn als Kandidaten zur Bemessung von Leistungswerten interessant machen und die nicht ausschließlich aus der Anzahl von Fließkommamultiplikationen pro Zeiteinheit bestehen. Neben den Multiplikationen werden auch Zwischenspeicherungen und Zugriffe auf den Hauptspeicher benötigt. Der Algorithmus der Transformation kann nicht in einzelne lineare Unterschritte aufgeteilt werden und muss somit innerhalb eines Funktionsaufrufs behandelt werden. Schließlich erhöht sich die Effizienz des Algorithmus mit zunehmender Blockgröße und die Kombination mehrerer Transformationen mit verschiedenen Blocklängen erlauben die Anpassung an verschiedene Problemstellungen der Signalverarbeitung in Echtzeit.

Es soll Software entwickelt und verwendet werden, deren Leistung dann auf verschiedenen Hardwareumgebungen untersucht werden soll. Neben den Abhängigkeiten der Leistung des Gesamtsystems von den Eigenschaften der

Hardwarearchitektur, des Betriebssystems und des Compilers bestehen noch weitere Einflüsse auf die Rechenleistung des Systems aus dem Bereich der Speicheradressierung und der Schleifenprogrammierung bei der Implementierung des Algorithmus (wie z. B. in Freed 1993 dargestellt). In Moorers⁵ (2000) wird eine Abschätzung zukünftiger Rechenleistung für die Audiosignalverarbeitung mit Standardcomputern ebenso aus den Erfahrungen mit der ‚schnellen Faltung‘ abgeleitet. Moorers errechnet einen jährlichen Zuwachs in der Verarbeitungsgeschwindigkeit um den Faktor 1.5 aus einer 22 Jahre alten Messung gegenüber einer aktuellen. Das Ergebnis entspricht in etwa dem Mooreschen Gesetz, wonach die Verdopplung der Rechenleistung ca. Alle 18 Monate erfolgt (siehe Abschnitt 5.4, Seite 56). Moorers Einschätzung der Extrapolation dieser Entwicklung ist:

„... Processes that were considered unfeasible because they were grotesquely complex will become matter-of-fact” (Moorers 2000).

Aus den Problemen, die aus der Verarbeitungsleistung erwachsen können, postuliert Moorers die Entwicklung von intelligenten Assistenten, die helfen sollen, die zukünftige Komplexität von Systemen zur Audioverarbeitung beherrschbar zu machen.

7.3.1 Schnelle Faltung als Testoperation

Um die Portabilität des Algorithmus auf verschiedene Hardware- und Betriebssysteme zu ermöglichen, wurde als Programmiersprache ANSI-C gewählt. Implementiert wurde der Algorithmus der diskreten Fourier-Transformation nach der Overlap-save Methode für reellwertige Signalfolgen und Filterkoeffizienten (Kammeyer und Korschel 1989).

In diesem Abschnitt soll der Aufwand für Transformation aus dem Zeitbereich in den Frequenzbereich, die Multiplikation im Frequenzbereich und die Rücktransformation in den Zeitbereich als zusammenhängende Operation betrachtet und als schnelle Faltung im Frequenzbereich bezeichnet werden. Der Aufwand der Hin- und Rücktransformation je Abtastwert ist im Allgemeinen proportional zum Logarithmus der Blocklänge und damit bei Blocklängen, die größer als etwa 1024 sind, deutlich effizienter als die direkte Filterung im Zeitbereich (Gardner 1995). Die zusätzliche Multiplikation im Frequenzbereich geht in den Aufwand als additive Größe ein.

⁵ Gemeint ist hier nicht Gordon Moore, sondern James A. Moorers.

7.3.2 Speichermodelle

Standardcomputersysteme verfügen über eine Hierarchie von Speichern für Prozessvariablen, um z. B. häufig verwendete Speicherplätze möglichst ‚nahe‘ am Prozessor abzulegen und damit den Zugriff auf den Speicher zeitlich optimal zu halten. Die verschiedenen Speicherebenen sind Prozessorregister, Cache-Speicher auf dem Prozessorchip, Hauptarbeitsspeicher des Computersystems und ausgelagerter Arbeitsspeicher auf der Festplatte. Um die verschiedenen Speichermodelle bei der Implementierung des Algorithmus zu untersuchen, wurde der Zugriff auf die Speichervariablen über Makros realisiert. Die Variablen bestehen aus Schleifenzählern, Werten für Zwischenspeicher und Zeigern auf Tabellen. Das eigentliche Argument der Faltungsroutine ist ein Zeiger auf einen zu transformierenden Signalblock. Die Variablen der Operation können definiert werden:

- als globale Variablen des Programms,
- als lokale Variablen der Routine oder
- als dynamische Variablen, die in eine Datenstruktur integriert sind.

In Bezug auf die verschiedenen Zielplattformen und Programmierumgebungen können dabei unterschiedliche Effekte entstehen. Üblicherweise sind die globalen Variablen statisch und bleiben auch nach der Beendigung der Routine im Hauptspeichersegment (*main memory segment*) des Programms existent. Im Gegensatz dazu werden bei dem Aufruf einer Routine alle lokalen Variablen auf einem sog. Stapel angelegt (*stack memory segment*) und nach Beendigung der Routine wieder freigegeben. Mit der in ANSI-C möglichen Speicherklassenangabe `register` ist es möglich, eine ‚Empfehlung‘ für die Behandlung einer Variablen während der Laufzeit zugeben, damit die Variable bevorzugt im Register des Prozessors gehalten wird und damit Speicherzugriffe optimiert werden. Der Speicherplatz der Variablenstruktur hingegen wird zur Laufzeit dynamisch im Arbeitsspeicher (*heap*) erzeugt und stellt somit eine Kombination aus den beiden vorhergehenden Speichermodellen dar. Damit soll die Möglichkeit gegeben werden, dass oft verwendeter Hauptspeicherplatz zusammenhängend vom Laufzeitsystem in den Zwischenspeicher (*cache*) auf dem Prozessor-Chip übertragen wird und dadurch schnellere Zugriffszeiten und Laufzeiten möglich werden. Unabhängig davon werden die Speicherbereiche für die Signalblöcke immer im dynamischen Arbeitsspeicherbereich erzeugt.

Die Makro-Funktion `FV()` wird verwendet, um zwischen den verschiedenen Speichermodellen vor der Übersetzung des Programms zu wählen. Über eine Makro-Variable wird diese Vorübersetzung gesteuert.

```
#ifdef FFT_GLOBAL  
    #define FV(ptr,var) ptr##_## var  
#endif  
#ifdef FFT_LOCAL  
    #define FV(ptr,var) var  
#endif  
#ifdef FFT_OBJECT  
    #define FV(ptr,var) ptr##_->## var  
#endif
```

Im Fall des globalen Speichermodells (FFT_GLOBAL) wird ein Symbol durch den Pre-Compiler erzeugt, über das nach der Übersetzung des Programms auf reservierten Speicherplatz im Hauptspeicher zugegriffen werden kann. Die lokalen Variablen (FFT_LOCAL) werden ohne weitere Spezifizierung am Beginn der Routine definiert. Der Zugriff auf dynamische Variablen (FFT_OBJECT) erfolgt über einen Zeiger auf den dynamischen Speicherplatz. Der Operator '##' bewirkt, dass nach der Vorkompilierung die Zeichenfolgen links und rechts vom Operator zu einer Zeichenfolge verbunden sind. Bei der Kompilierung wird diese Zeichenfolge zu einem Symbol verarbeitet.

In den folgenden Graphen der Testmessungen sind die Leistungswerte der Verarbeitung mit dem Algorithmus über einer logarithmischen X-Achse aufgetragen, d. h. die Blockgröße wird linear mit dem Exponenten der Basis 2 aufgetragen. Die Abkürzung 'K' steht bei den Blocklängen nicht für den Faktor Tausend, sondern für 1024, d. h. 2^{10} . Die Angabe auf der Y-Achse gibt die Verarbeitungsdauer der Testoperation bezogen auf einen Abtastwert des Signalblocks an. Damit werden die Angaben zum Aufwand der Operation unabhängig von dem Prozessortyp und es können verschiedene Systeme miteinander verglichen werden. Es ergibt sich theoretisch im Graphen ein linearer Verlauf des Zusammenhangs zwischen Blockgröße und Zeitaufwand der Berechnung pro Abtastwert. Wie aber zu sehen sein wird, verhält sich dieser Zusammenhang unter realen Bedingungen abweichend von einer Geraden. Die Gründe dafür sind in den Bereichen der Prozessorarchitektur, des zur Übersetzung verwendeten Compilers und des Betriebssystems zu finden. Stellenweise sind horizontale Linien in die Graphen eingetragen, die angeben, bis zu welcher Blockgröße die Verarbeitung von Monosignalen in Echtzeit bei einer bestimmten Signalabtastrate möglich ist. Bei Leistungswerten oberhalb der Grenze dauert eine Verarbeitung länger als die Zeitspanne einer entsprechenden Anzahl von Abtastwerten. Für ein einkanaliges Audiosignal der

Abtastrate 44.1 kHz beträgt z. B. diese Grenze 22.67 μ s und für ein Stereosignal 11.34 μ s. In den Graphen wird zum Teil auch die 4- bzw. 8-fach Abtastrate von 44.1 kHz berücksichtigt.

Drei Eigenschaften werden bei der Ausführung der Testoperation auf verschiedenen Standardcomputersystemen deutlich:

- im Bereich der Blocklänge von 2^{11} , d. h. 2K Abtastwerten kommt es bei der Verarbeitung auf bestimmten Prozessortypen zu einer signifikanten, stufenweisen Erhöhung des Berechnungsaufwands (siehe Abb. 37),
- ab einer Blockgröße von 2^{15} bzw. 2^{16} , d. h. 32K bis 64K Abtastwerte, steigt der Berechnungsaufwand der Faltung teilweise unverhältnismäßig stark an, so dass es günstiger erscheint Operationen oberhalb dieser Blocklänge durch mehrere Operationen mit kürzerer Blocklänge auszuführen (siehe Abb. 38),
- der Verlauf des Berechnungsaufwands über der Blocklänge verhält sich bei den getesteten RISC Prozessorsystemen nahezu linear. Die beiden oberen Merkmale sind bei diesen Systemen kaum oder gar nicht ausgeprägt (siehe Abb. 39).

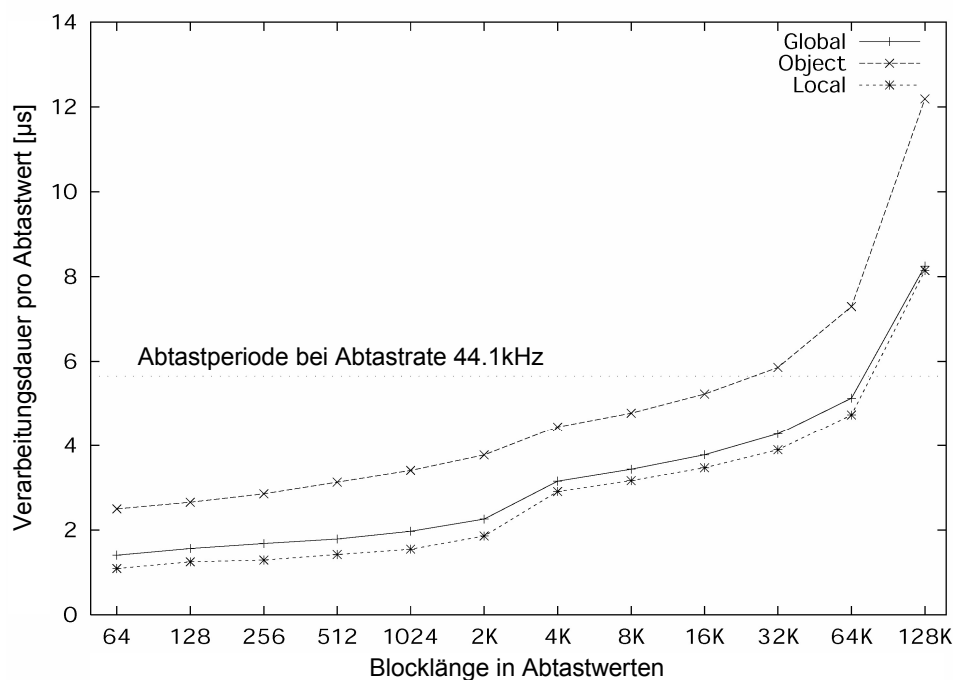


Abb. 37: Verarbeitungswerte eines Standardprozessors (350 MHz) bei verschiedenen Blocklängen: Wird das Speichermodell ‚object‘ verwendet, wird knapp doppelt soviel Verarbeitungsleistung erfordert als mit dem Speichermodell ‚local‘ bzw. ‚global‘.

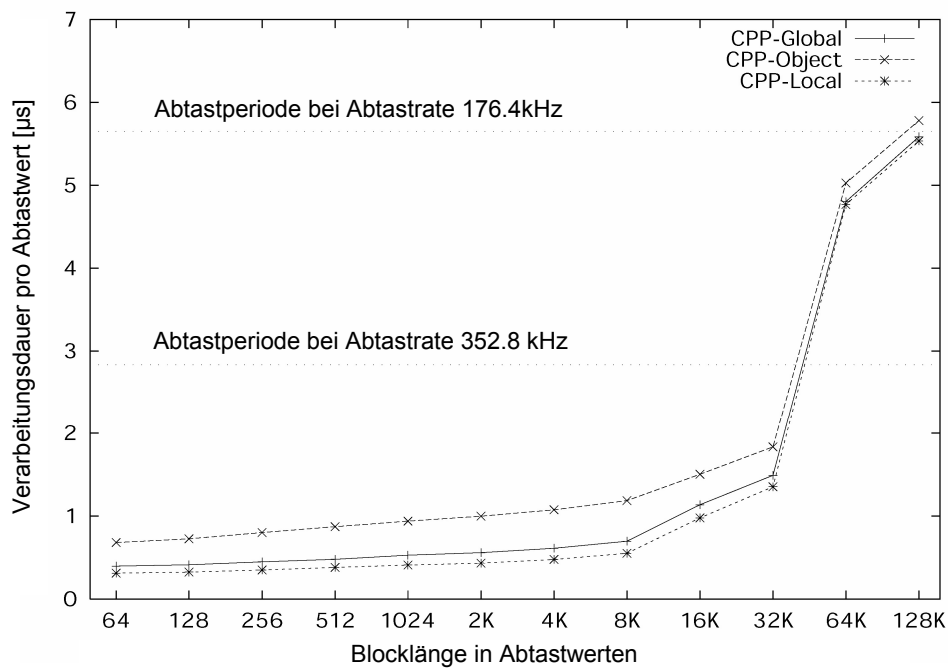


Abb. 38: Verarbeitungswerte eines Standardprozessors (1.2 GHz) bei verschiedenen Blocklängen: Der lineare Verlauf der Messkurve ist bis zu einer Blockgröße von 2^{15} ist typisch für moderne Prozessoren.

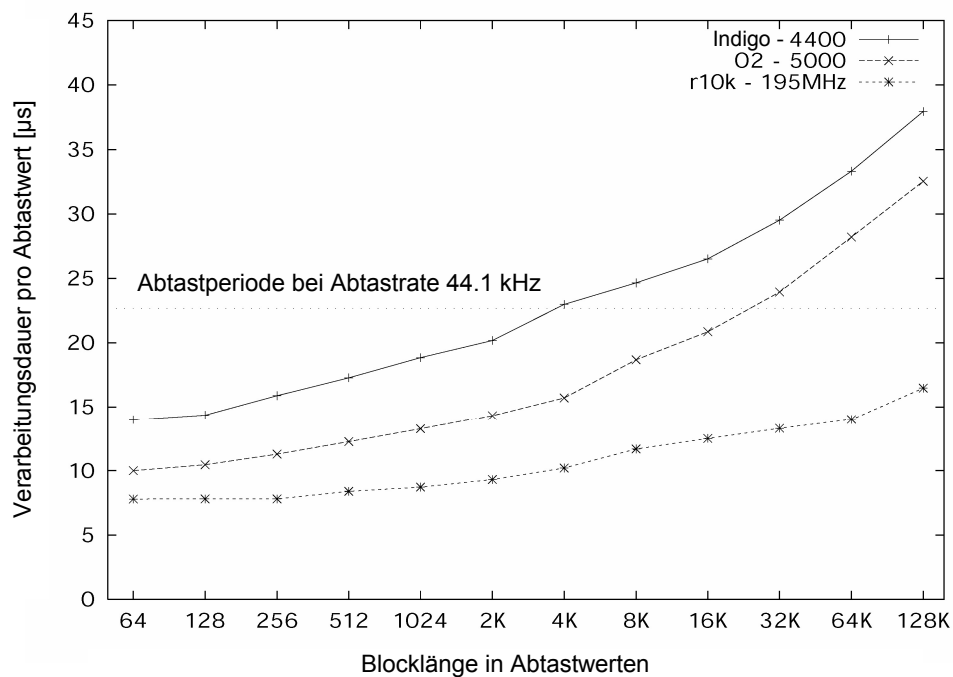


Abb. 39: Die Leistungswerte der schnellen Faltung auf RISC-Prozessorsystemen mit unterschiedlichen Taktraten: Der Verlauf der Zunahme der Bearbeitungszeit ist nahezu linear.

7.3.3 Hardware-Plattformen und Betriebssysteme

Die Computersysteme, für die die Bewertung der Leistungsfähigkeit und der Eigenschaften der verschiedenen Einflussgrößen untersucht wurden, waren hauptsächlich Computer mit dem Betriebssystem Windows. Untersuchungen an RISC Computern wie Silicon Graphics wurden anfänglich gemacht, aber aufgrund der immer größeren Verfügbarkeit von kostengünstigen Standardcomputern bald aufgegeben. Als Betriebssystem wurde Windows NT und Windows 98 sowie BeOS zur Bewertung herangezogen. Lediglich das Betriebssystem BeOS verfügt über die Möglichkeit der Zuordnung von Echtzeitprozessen mit definierten Prioritäten, die nicht durch Prozesse mit niedriger Priorität unterbrochen werden können. Dagegen stehen unter Windows lediglich Prioritätsklassen zur Auswahl, innerhalb derer Prioritätsabstufungen gemacht werden können. Der Name der höchsten Prioritätsklasse `REALTIME_PRIORITY_CLASS` weist fälschlicherweise auf die Möglichkeit der Echtzeitfähigkeit hin. Die Klasse hat aber lediglich die höchste Priorität und kann dennoch durch Prozesse mit niedrigerer Priorität unterbrochen werden. Das Windows Betriebssystem arbeitet im Allgemeinen nach dem Prinzip einer Zeitscheibe, wonach jeder Prozess nacheinander für eine bestimmte Zeit ausgeführt wird. Eine Ausführung des Prozesses zu einem bestimmten Zeitpunkt ist jedoch nicht garantiert. Besteht allerdings ausreichend Rechenleistung und wird auf die Ausführung anderer aufwändiger Prozesse durch den Benutzer verzichtet, wird ein Echtzeitverhalten annähernd möglich.

Das Betriebssystem BeOS verfügt über die Möglichkeit neben der Zeitscheibenverwaltung von Prozessen zusätzlich Echtzeitprozesse zu definieren. In einem speziellen Bereich der Werte der Prioritäten können Prozesse ausgeführt werden, die nur durch Echtzeit-Prozesse mit höherer Priorität unterbrochen werden können. Damit werden die Zeitpunkte der Ausführung garantiert und vorhersagbar. Die Programmierung von derartigen Echtzeitsystemen erfordert jedoch besondere Aufmerksamkeit, da ein falsch programmierter Prozess das gesamte Computersystem blockieren kann. Die Echtzeiteigenschaft des Betriebssystems erfordert aber auch einen zusätzlichen Aufwand innerhalb der Prozessverwaltung des Betriebssystems, was sich in den leicht höheren Leitungskurven von BeOS gegenüber Windows 98 ausdrückt (siehe Abb. 40).

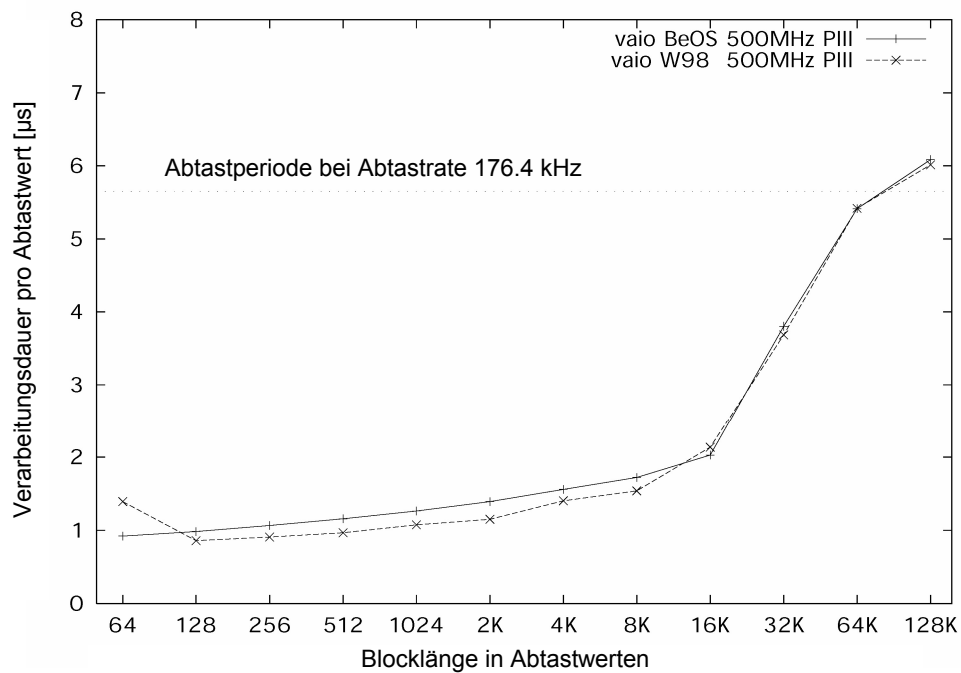


Abb. 40: Wird die ‚schnelle Faltung‘ auf demselben Prozessor mit unterschiedlichem Betriebssystem ausgeführt, kann das Echtzeitbetriebssystem am höheren Aufwand erkannt werden.

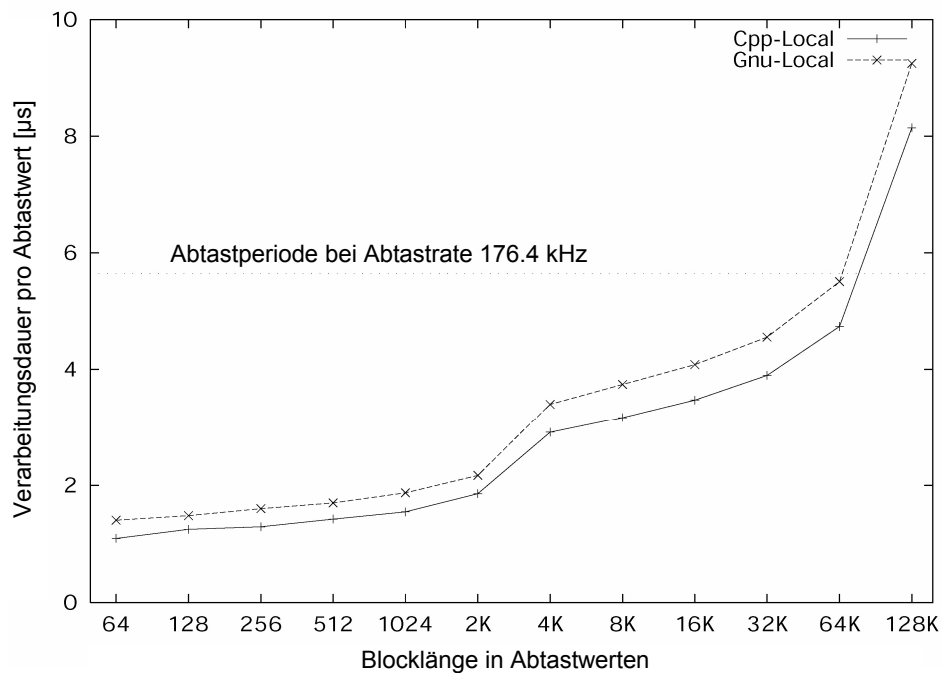


Abb. 41: Das Werkzeug zur Übersetzung des Programmcodes zur ‚schnellen Faltung‘ im Frequenzbereich hat einen deutlichen Einfluss auf die Verarbeitungsgeschwindigkeit.

Dieser Effekt könnte aber auch vom unterschiedlichen Compiler herrühren. Dies wird besonders dann deutlich, wenn der gleiche Quellcode mit unterschiedlichen Compilern übersetzt wird und die resultierenden Programme auf demselben Computer in ihrer Leistungsfähigkeit untersucht werden. So z. B. ist ein durch den Microsoft C++ Compiler erzeugtes Programm in der Verarbeitung um ca. 17% schneller als das von einem GNU-Compiler erzeugte Programm (siehe Abb. 41).

7.3.4 Prozessorspezifische Optimierung

Um eine Abschätzung für den Leistungszuwachs mit einer prozessorspezifischen Optimierung für den Algorithmus der schnellen Faltung zu erhalten, wurde der ANSI-C Code der Implementierung an geeigneten Stellen um sog. Assembler-Inlines modifiziert. Damit ist es möglich, an bestimmten Stellen im Code explizite Anweisungen in Maschinensprache einzufügen, um die besonderen Befehls- und Verarbeitungsmöglichkeiten des spezifischen Prozessors auszunutzen. Andere Optimierungsmethoden für die FFT, wie in Frigo und Johnson (1998) beschrieben, basieren auf adaptiven Verfahren, bei dem der verwendete Code dem speziellen Prozessor angepasst ist. Ein sog. Plan der Verarbeitung wird vor der Ausführung des Programms erstellt und kann sich somit verschiedenen Zielsystemen anpassen. Für diese Arbeit soll jedoch lediglich die Möglichkeit moderner Prozessoren ausgenutzt werden, im sog. pipelined Verfahren mehrere Operationen auf Daten mit einem Befehl in einem Prozessortakt auszuführen (*Single Instruction Multiple Data*, SIMD). Der so optimierte Algorithmus soll dann in seinem Verhalten mit der reinen portierbaren C-Implementierung verglichen werden. An geeigneten Stellen der Schleifen innerhalb der Implementierung der schnellen Fourier-Transformation wurden Multiplikationen und Additionen von komplexen Wertepaaren durch den speziellen Befehlssatz des Prozessors beschleunigt. Leistungsmessungen ergaben im Schnitt, dass bis zu einer Blocklängenverarbeitung von 2^{14} (16K) Abtastwerten die Berechnungszeit für die schnelle Faltung durch die Assembler Optimierung nur ca. 66% der Berechnungszeit der reinen C-Implementierung beträgt. Oberhalb der Blocklänge von 2^{14} Abtastwerten ergeben sich keine derart deutlichen Verbesserungen mehr, was darauf zurückzuführen ist, dass bei dem verwendeten Prozessortyp (AMD-Athlon) die Größe des primären Datenpuffers von 64 KBytes überschritten wurde.

7.3.5 Ergebnisse der Leistungsvergleiche

Der zentrale Parameter der Leistungsfähigkeit eines Standardcomputersystems für die Operation der schnellen Faltung ist die Taktrate des Prozessors (siehe Abb. 42). Zusätzlichen Einfluss auf den Verlauf der Verarbeitungsdauer über der Blocklänge hat

der Prozessortyp und das verwendete Betriebssystem. Echtzeitbetriebssysteme bzw. RSIC Prozessorsysteme oder neuere Prozessoren haben eine Leistungskurve, die weniger sprunghafte Änderungen aufweist als die der anderen getesteten Systeme.

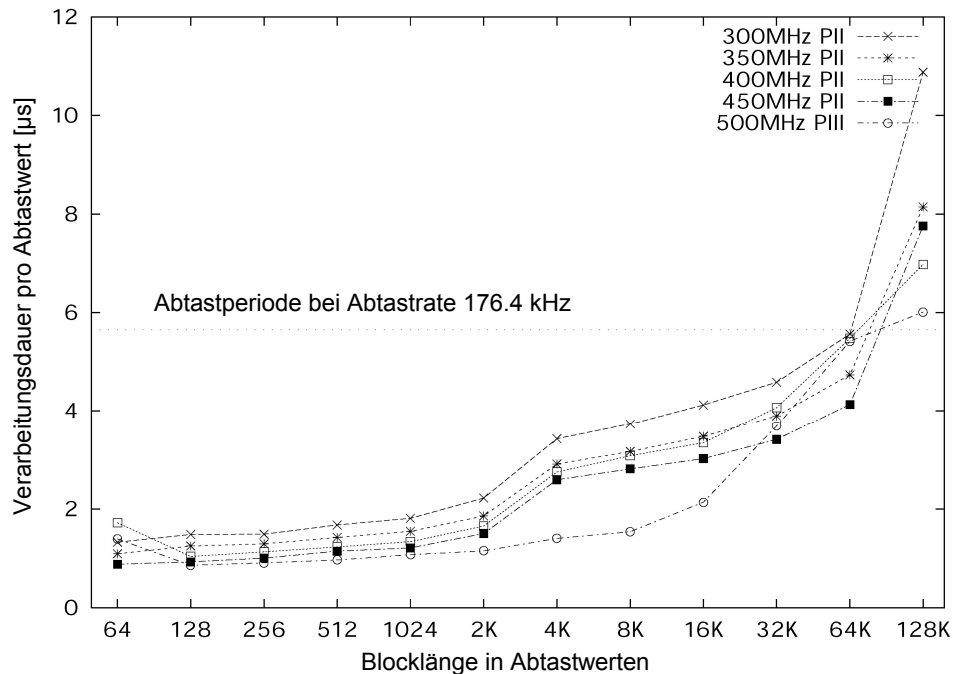


Abb. 42: Die Taktrate des Prozessors, auf dem die Berechnung erfolgt, ist signifikant für die Geschwindigkeit, mit der der Algorithmus ausgeführt wird. Weitere Unterschiede ergeben sich aus dem Verlauf der Leistungskurve.

Bei dem verwendeten Echtzeitbetriebssystem wird dieses Verhalten jedoch durch einen geringen zusätzlichen Aufwand bei der Prozessverwaltung gewonnen (siehe Abb. 40). Möglicherweise ist der zusätzliche Aufwand aber auch auf eine schlechte Übersetzung des Programms zurückzuführen. Der verwendete Compiler mit den jeweiligen Optionen zur Optimierung des Laufzeitverhaltens hat zusätzlichen, aber geringeren Einfluss auf die Verarbeitungsgeschwindigkeit (siehe Abb. 41). Die Leistungswerte verlaufen nicht ideal und es gilt daher immer zu beachten, dass ein erwarteter Gewinn an Leistungskapazität sich nicht zwangsläufig einstellt, wenn die Verarbeitung mit einer größeren Blocklänge erfolgt. Eine sog. optimale Kennlinie der Leistung ergibt sich, wenn berücksichtigt wird, dass an Stellen sprunghaften Anstiegs des Verarbeitungsaufwands mit geringerer Blockgröße bessere Werte erreicht werden (so z. B. in Abb. 38, Seite 125) ab den Werten für $n=2^{15}$, 32K und $n=2^{16}$, 64K).

Durch eine Optimierung der Implementierung für den jeweiligen Prozessortyp z. B. unter Verwendung von SIMD Befehlen kann sich ein deutlicher Leistungsgewinn bei der Verarbeitung einstellen, der jedoch zulasten der allgemeinen Portierbarkeit geht.

7.4 Signalverarbeitung in parallelen Prozessen

Das Beispiel der schnellen Faltung aus dem vorherigen Abschnitt wird dazu herangezogen, um die Problematik der Komplexität und Ausgabeverzögerung dieser Operation unter der Echtzeitbedingung darzustellen. Mithilfe der parallelen Programmierung auf Standardcomputersystemen wird dazu eine portierbare Lösung erarbeitet, die der Anforderung interaktiver Anwendungen genügt.

Auch wenn auf der Benutzungsoberfläche von Anwendungen die Signalströme als parallel verarbeitet dargestellt werden (wie z. B. in Max/ MSP 1997 oder Pure Data 2001), so erfolgt eine zugrundeliegende Blocksignalverarbeitung nacheinander. Der Verarbeitungsaufwand pro Block ist in der Regel konstant und linear teilbar, so dass z. B. eine Verarbeitung mit der halben Blocklänge auch zu einer Halbierung des Verarbeitungsaufwands führt. Für viele Anwendungen kann von diesen Annahmen ausgegangen werden. Wie jedoch im vorherigen Abschnitt zu sehen war, gibt es auch Berechnungsverfahren, die im Aufwand nicht linear mit der Blockgröße steigen, wobei gleichzeitig die Berechnungsvorschrift schlecht teilbar ist. In solchen Fällen kann eine parallele Signalverarbeitung auch auf der Ausführungsebene der Anwendung dazu verwendet werden, um eine einfache Realisierung der Verarbeitung zu ermöglichen. Dazu werden z. B. die Möglichkeiten des Betriebssystems zur parallelen Verarbeitung genutzt.

Heutige Standardcomputerumgebungen bieten die Möglichkeit, mehrere Prozesse auf einem Prozessor gleichzeitig auszuführen und darüber hinaus Mittel zur Kommunikation zwischen den Prozessen, damit eine Synchronisation zwischen ihnen möglich ist. Mit einem Algorithmus zur effizienten Faltung von Signalen mit langen Impulsantworten soll demonstriert werden, wie mithilfe von Standardcomputern leistungsaufwändige Verfahren der Signalverarbeitung effizient realisiert werden können und gleichzeitig eine geringe Ausgabeverzögerungszeit des Gesamtsystems zu erhalten. Die zuletzt genannte Eigenschaft ist gerade eine zentrale Forderung für die Interaktion eines Benutzers mit einer akustischen Darstellung in Echtzeit (siehe auch Abschnitt 2.2.1, Seite 11).

Im Bereich der Audiosignalverarbeitung und speziell in der dreidimensionalen Klangprojektion (Begault 1994) benötigen die Anwendungen oft eine Faltungsoperation, die in Echtzeit ausgeführt wird und sehr viele Koeffizienten besitzen kann (Kleiner et al. 1993). Eine Anzahl von Koeffizienten von 1024 und mehr ist keine Seltenheit. Im Kontext von virtueller Raumakustik und binauraler Präsentation bedeutet dies, dass die Filterkoeffizienten des linearen und zeitinvarianten Systems (linear time-

invariant, LTI) einem Eingangssignal zusätzliche Informationen aufprägen, wie z. B. die Nachhallcharakteristik eines Raums oder die Frequenzcharakteristik der Schalleinfallrichtung. Die Filterung durch ein System mit einer endlichen Impulsantwort (finite impulse response, FIR) wird als Faltung bezeichnet und die äquivalente Operation im Frequenzbereich ist die Multiplikation (Fliege 1991). Die Faltung im Zeitbereich benötigt einen hohen Aufwand pro Abtastwert, der theoretisch proportional mit der Anzahl der Filterkoeffizienten N steigt. Für Echtzeitanwendungen benötigt diese Art der Faltung eine sehr hohe Rechenleistung. Die Verarbeitung der Faltung im Frequenzbereich kann dagegen sehr effizient realisiert werden und der Aufwand pro Abtastwert ist lediglich proportional dem Logarithmus der Filterlänge (Kammeyer und Kroschel 1989). Jedoch erscheint hierbei das Problem einer deutlich höheren Ausgabeverzögerung, die durch die Blockverarbeitung im System bedingt ist (siehe Abschnitt 4.2.1, Seite 43).

Das dadurch entstehende Dilemma der Faltungsoperation für Echtzeitanwendungen besteht darin, nicht gleichzeitig eine effiziente Verarbeitung bei niedriger Ausgabeverzögerung zu erhalten. Darüber hinaus ist eine Verminderung der Ausgabeverzögerung nicht prinzipiell durch ein leistungsstärkeres Rechnersystem zu erzielen, sondern ist inhärent im implementierten Algorithmus enthalten (insbesondere auch durch die Blockverarbeitung). Im Zusammenhang mit Echtzeit-Audioanwendungen ist eine geringe Ausgabeverzögerung immer eine angestrebte Eigenschaft. Zu hohe Ausgabeverzögerungen können vom Benutzer als Echo wahrgenommen werden, bei der Interaktion zu Irritationen führen oder die Illusion einer virtuellen akustischen Umgebung beeinträchtigen (siehe auch Abschnitt 3.1.3, Seite 29).

7.4.1 Beispiel einer Implementierung eines Faltungsalgorithmus

Der Algorithmus zur effizienten Faltung in Echtzeit, bei gleichzeitiger Realisierung geringer Verzögerungszeiten zwischen Ein- und Ausgabe der Signaldaten, besteht im Wesentlichen darin, dass das Gesamtproblem in Subprobleme aufgeteilt wird, die je einen Teil der Filterantwort verarbeiten (siehe Abb. 43). Dazu wird die Filterimpulsantwort in Blöcke unterschiedlicher Länge zerlegt. Grundlegende Arbeiten in diesem Bereich wurden von Stockham (1966) und Egelmeers und Sommen (1994) gemacht. Gardner (1995) realisierte den Algorithmus auf einem DSP-System und verwendete das System zur Verhallung von Audiosignalen. Ein kommerzielles Produkt wurde um 1995 von der Firma lakedsp entwickelt (Huron 2000) und basierte auf einer speziellen Hardware, die auf die oben beschriebene Verarbeitung zugeschnitten ist.

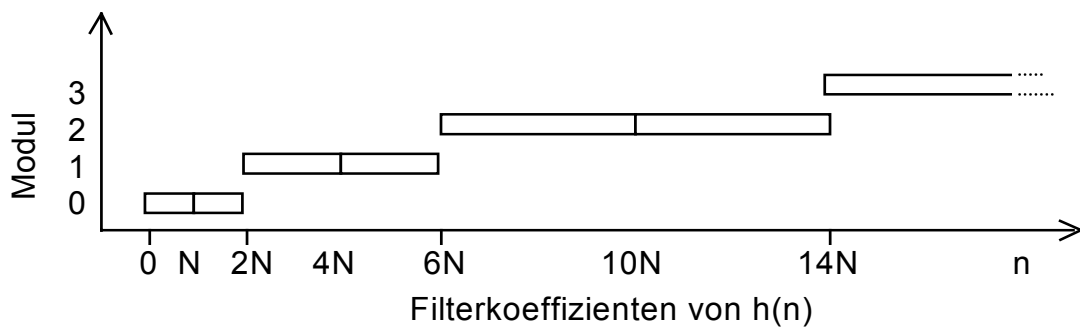


Abb. 43: Diagramm der Dekomposition einer Impulsantwort $h(n)$ zur Verarbeitung in parallelen Prozessen. Die Organisation erfolgt in ungleichen Blocklängen.

7.4.2 Mechanismus der Verarbeitung

Die Abtastwerte des Eingangssignals werden zuerst in Blöcke gruppiert, um die Faltung im Frequenzbereich mithilfe der schnellen Fourier-Transformation in parallelen Modulen zu ermöglichen. Dabei durchläuft jedes der parallel arbeitenden Module drei Stadien bei der Verarbeitung (siehe Abb. 44). Im ersten Stadium werden die eintreffenden Abtastwerte in den Signalblock aufgefüllt. Dann wird der Signalblock der eigentlichen Verarbeitung unterzogen, d. h. die Faltung wird im Frequenzbereich durchgeführt. Im dritten Zustand wird dann der transformierte Block für die Ausgabe ausgelesen (siehe auch Abschnitt 4.2.1, Seite 43).

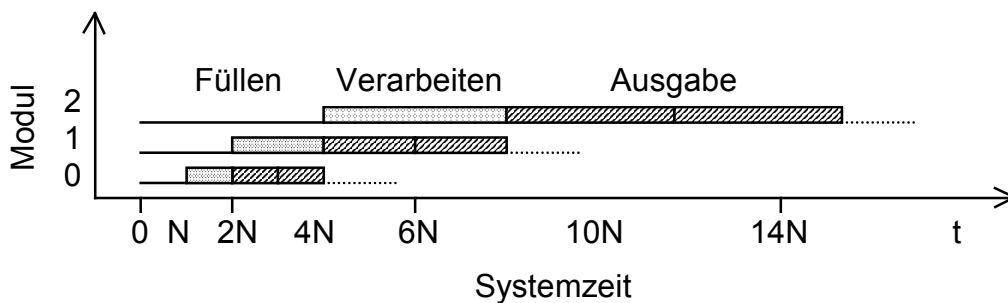


Abb. 44: Die verschiedenen Stadien der Verarbeitung der parallel arbeitenden Module. Die Ausgabeverzögerung eines Moduls wird durch die Ausgabeblöcke der vorherigen Module ausgefüllt.

Unter der Bedingung einer Echtzeit-Anwendung besteht somit eine zeitliche Verzögerung in der Systemantwort eines Moduls aus der Zeit, die benötigt wird, um einen Signalblock zu füllen, zuzüglich einer Zeitdauer gleicher Größe, in der die Faltung ausgeführt wird. Die resultierende Verzögerung ergibt sich zu:

$$D_m = 2 \frac{N_m}{f_s}, \quad (8)$$

wobei f_s die Abtastrate und N_m die Anzahl der Abtastwerte in einem Block des Moduls m beschreibt. Die Aufteilung der Filterimpulsantwort, wie sie in Abb. 43 dargestellt ist, hat für die Verzögerungszeit des Gesamtsystems aller Module die folgende Konsequenz: Die Verzögerung von Modul 0 beläuft sich auf $2N$ und somit steht der erste Ausgabeblock nach $2N$ Abtastwerten zur Verfügung. Die Ausgabe des Moduls dauert bis zum Zeitpunkt $t = 4N$ (siehe Abb. 44), da die Filterlänge $2N$ beträgt. Parallel dazu startet das Modul 1 zum Zeitpunkt $t=0$, um seinen Eingangssignalblock zu füllen. Die Gesamtdauer der Ausgabeverzögerung von Modul 1 beträgt daher $4N$. Dies ist exakt der Zeitraum, der durch die Verzögerung und die Ausgabe von Modul 0 gedeckt wird. Die induktive Betrachtung kann allgemein so formuliert werden:

Die Ausgabeverzögerung eines Moduls m ist so lang wie die Ausgabeverzögerung und die Ausgabedauer des vorherigen Moduls $m-1$.

Die Aufteilung der Impulsantwort des Filters in den Modulen mit unterschiedlichen Blocklängen garantiert, dass keine Lücke in der Ausgabe der Verarbeitung erscheint. Die Gesamtausgabeverzögerung des Systems ist die des Moduls 0 (d. h. $2N_0$) und also unabhängig davon, wie viele Module höherer Ordnung parallel arbeiten. Die Länge L eines Filters mit mehreren Modulen ergibt sich aus der Summe der Längen der einzelnen Module:

$$L(N_0, M) = \sum_{m=0}^{M-1} 2N_m = \sum_{m=0}^{M-1} N_0 2^{m+1}. \quad (9)$$

M bezeichnet die Anzahl der Module und N_0 bestimmt die Anzahl der Abtastwerte des Blocks von Modul $m=0$. Die Anzahl der Filterkoeffizienten ist daher nicht frei wählbar, sondern erfolgt immer gemäß der Gleichung 9. Die Frequenz f_m , mit der in einem Modul die Verarbeitung erfolgt, ergibt sich aus der Dauer eines Blocks von N_m Abtastwerten:

$$f_m = \frac{f_s}{N_m} \text{ oder } T_m = \frac{N_m}{f_s}. \quad (10)$$

Üblicherweise liegt die Blocklänge des ersten Moduls in der Größe von 64 bis 1024 Abtastwerten. Damit ergibt sich eine Blockverarbeitungsfrequenz von 690 Hz bis zu 43 Hz bei einer Abtastrate von 44.1 kHz. Dies entspricht üblichen Werten für interaktive Anwendungen. Module höherer Ordnung können eine Blockverarbeitungsfrequenz von

unter einem Hz besitzen. Diese besondere Eigenschaft kann z. B. benutzt werden, wenn frühe Anteile des Filters häufiger gewechselt werden müssen als späte (z. B. bei Raumimpulsantworten, Müller-Tomfelde 1999 und 2001).

Die Signalverarbeitung innerhalb der Module besteht im Wesentlichen aus den Operationen der schnellen Fourier-Transformation und der inversen Operationen dazu (inverse fast Fourier transformation, IFFT). Im Frequenzbereich erfolgt noch die elementweise Multiplikation des transformierten Eingangssignalblocks mit der transformierten Filterimpulsantwort (Kammeyer und Kroschel 1989, Egelmeers und Sommen 1994).

7.4.3 Gesamtsystem der Verarbeitung

Im Kern besteht das gesamte Verarbeitungssystem aus M Modulen, die parallel arbeiten. Die eingehenden Abtastwerte werden in Blöcke zusammengefasst bzw. der serielle Datenfluss wird in einen blockweise parallelen Datenfluss konvertiert. Die Blöcke werden dann gemäß der benötigten Blocklänge N_m an die verschiedenen Module weitergeleitet. Die Verwaltung der internen Signalblöcke und deren Anbindung an die Ein- und Ausgabe erfolgen über ‚schaltbare‘ Speicherobjekte, die mit Signalblöcken der Länge N_0 gefüllt und ausgelesen werden. Die Verarbeitung erfolgt in jedem Modul dann in der Länge N_m . Diese Speicherverwaltung entkoppelt die Länge der Eingangs- und Ausgangssignalblöcke N_0 von der Länge der modulinternen Blöcke N_m . Die Blockgröße der jeweils folgenden Module verdoppelt sich jeweils, so dass gilt: $N_{m+1} = 2N_m$. Die Blockverteilungseinheit versorgt jedes Modul zum geeigneten Zeitpunkt mit Signalblöcken der benötigten Blocklänge. Der Speicherbedarf der Verarbeitung kann stark ansteigen. So beträgt z. B. der benötigte Speicher für Signalblöcke bei einer Faltung mit einer Impulsantwort von ca. 1.4 sec. mit einem System mit vier Modulen ca. 500 KByte und kann bei längeren Impulsantworten bis zu mehreren MBytes reichen. Dieser Speicher wird während der Verarbeitung ständig beschrieben und ausgelesen.

Das System wie in Abb. 46 dargestellt und der Algorithmus kann mit geringen Modifikationen so erweitert werden, dass Teile der Koeffizienten $h(n)$ während der Verarbeitung in Echtzeit ohne erheblichen Mehraufwand ausgetauscht werden können. Das Gesamtsystem wird dadurch ein System mit linearen zeitveränderlichen Eigenschaften (linear time-variant, LTV). Für die Koordination des Austauschs der Koeffizienten aller Module sind verschiedene Strategien möglich, die je nach Anwendung verwendet werden können. Sollen nur wenige zusammenhängende Koeffizienten in der Impulsantwort ausgetauscht werden, beschränkt sich die Operation lediglich auf das Modul, in dem diese Koeffizienten benötigt werden. Für die

Verwendung in interaktiven Anwendungen kann auch die Eigenschaft des Systems ausgenutzt werden, dass frühe und kurze Anteile in der Impulsantwort häufiger gewechselt werden können als späte und große Anteile (Müller-Tomfelde 2001).

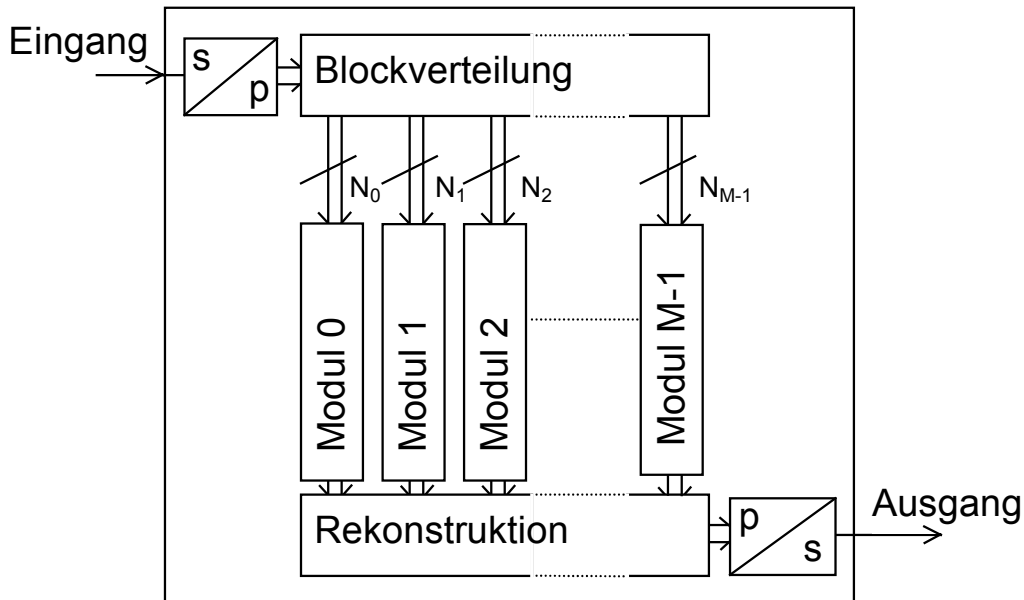


Abb. 46: Die Architektur des Verarbeitungssystems mit M parallel operierenden Modulen, die je einen Teil der Gesamtfaltung mit $2N_m$ Koeffizienten ausführen.

Die Signalblöcke am Ausgang der Module werden schließlich so addiert, dass sich die Impulsantwort des Gesamtsystems einstellt. Man kann hier auch von einer Rekonstruktion im Zeitbereich sprechen, da das System im statischen Fall von außen betrachtet eine Faltung mit der Impulsantwort $h(n)$ vollzieht.

7.4.4 Ergebnisse der Implementierung

Für den oben beschriebenen Algorithmus wurde eine vollständig portierbare Softwarelösung geschaffen, die auf Standardcomputern zur Ausführung gebracht werden kann. Der Aufwand der Entwicklung liegt nicht im Bereich der Verarbeitung mit der schnellen Fourier-Transformation innerhalb der Module, sondern vielmehr in der zusätzlichen Steuerung und Verwaltung der parallelen Prozesse. Über eine Funktionsschnittstelle sind Zugriffe auf geschützte Variablen und Signale zur Interprozesskommunikation (*semaphore*) unter verschiedenen Betriebssystemen realisiert (Windows, NeXTStep, BeOS).

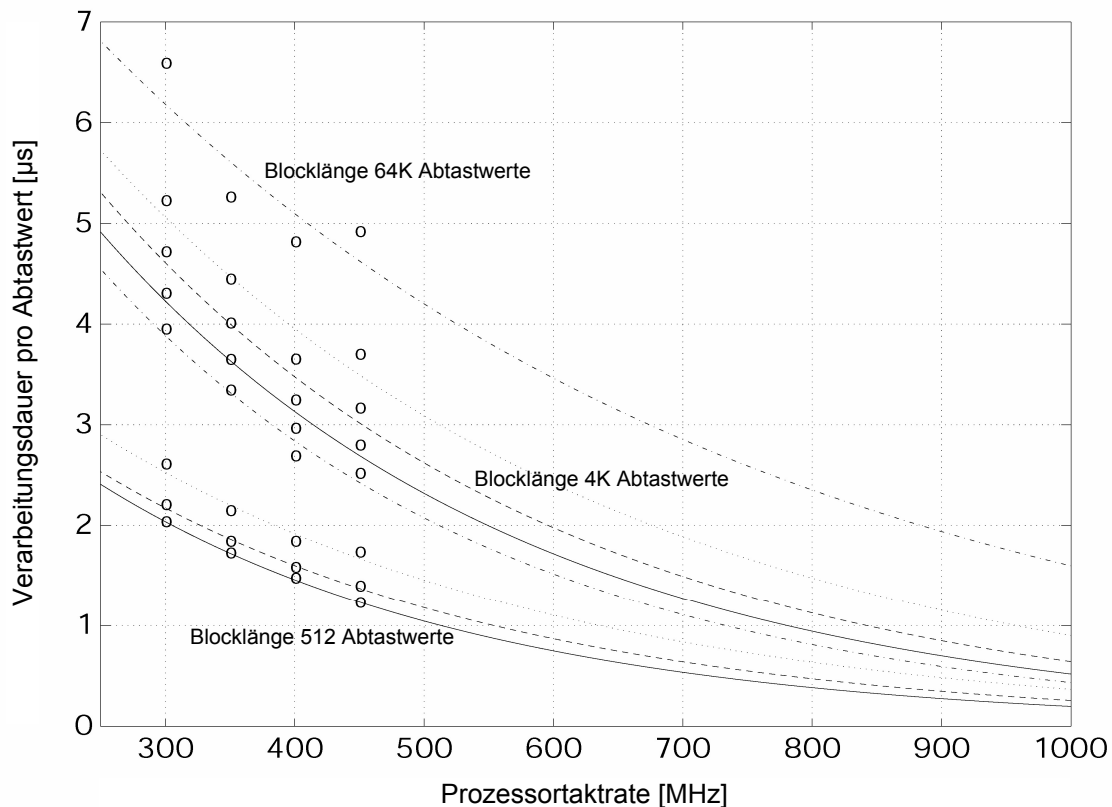


Abb. 47: Extrapolation des Aufwands der Faltung im Frequenzbereich mithilfe der Fourier-Transformation für drei unterschiedlich große Blocklängen.

Diese Funktionen des jeweiligen Betriebssystems sind notwendig, um eine termingerechte Ausführung der Prozesse und die Synchronisation der Module in Echtzeit zu ermöglichen. Da die Prozesse verschiedene Ausführungsgeschwindigkeiten besitzen, wurden ihnen auch verschiedene Prioritäten zugeordnet. Unter dem Echtzeitbetriebssystem BeOS wurde somit ein ratenmonotones Planungsverfahren für die Verarbeitung der Prozesse angewendet (Steinmetz 2000). In einer zentralen Funktion der Implementierung werden die auf ihre Aktivierung wartenden Module periodisch gestartet. Sollte der Prozessor überlastet sein, kann es hierbei auch zu einem Überlauf der Prozesse kommen.

Der zusätzliche Aufwand der Prozess- und Speicherverwaltung macht sich im Gesamtaufwand der Ausführung nur minimal bemerkbar. Kern der Verarbeitung bleiben die Leistung verbrauchenden Transformationen innerhalb der Module. Durch die Implementierung parallel verarbeitender Prozesse wird jedoch abhängig vom Betriebssystem eine Instabilität bei der Verarbeitung möglich, so dass immer auch ein Reserve (*headroom*) an Verarbeitungsleistung eingeplant werden muss.

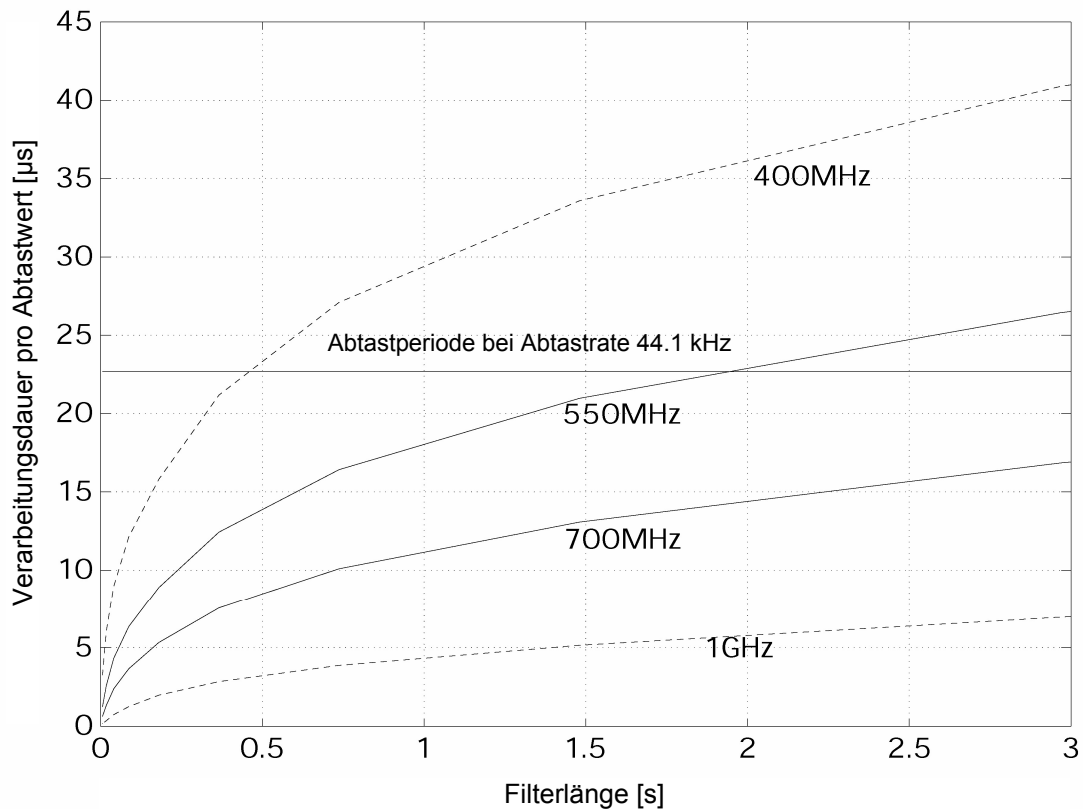


Abb. 48: Ergebnis der Berechnung des Aufwandes des Faltungsalgorithmus mit geringer Ausgabeverzögerung auf der Grundlage der Werte der extrapolierten Prozessorleistungen. Die Verarbeitungsdauer ist als Funktion der Länge des Filters dargestellt. Die Systemblocklänge beträgt $N_0=256$ Abtastwerte.

Auf der Grundlage von Testmessungen auf verfügbaren Computersystemen wurden Extrapolationen der Verarbeitungszeiten der Module von Prozessoren höherer Taktraten berechnet (siehe Abb. 47), die zum Zeitpunkt der Messung noch nicht verfügbar waren. Beim Vergleich des Aufwands aus der Abschätzung (Abb. 47) mit dem einer aktuellen Messung eines 1.2 GHz Prozessors (siehe Abb. 38, Seite 125) wird die Richtigkeit der Extrapolation erkennbar: In beiden Fällen liegt der Aufwand bei kleinen Blocklängen unterhalb von ca. 1 μ s.

Aus den Werten dieser extrapolierten Verarbeitungszeiten konnte der Aufwand der Faltung mit paralleler Signalverarbeitung bei geringer Ausgabeverzögerung vorherberechnet und abgeschätzt werden. Die Verarbeitungsdauer einer derartigen Faltung bei verschiedenen Prozessortaktraten ist in Abb. 48 als Funktion der Filterlänge dargestellt. Die Ausgabeverzögerung der Operation beträgt ca. 11.61 ms. Aus den extrapolierten Werten geht hervor, dass ab einer Taktrate von ca. 700 MHz eine

Echtzeitverarbeitung möglich ist. Die bedeutet, dass dann z. B. eine Verarbeitung eines Monosignals bei der Abtastrate 44.1 kHz mit einer Filterlänge bis zu drei Sekunden mit dem oben beschriebenen Algorithmus möglich ist. Es konnte damit gezeigt werden, dass einer solchen komplexen Verarbeitung mit Standardcomputern in nächster Zukunft nichts im Wege steht. Die Übertragung der Implementierung auf ein Standardcomputersystem mit einer Prozessortaktrate oberhalb der 1 GHz Frequenz bestätigt dieses Ergebnis.

8 Zusammenfassung

Die Ergebnisse dieser Arbeit lassen sich wie folgt zusammenfassen: Ausgehend von einer Analyse der Mensch-Computer Interaktion bei veränderten technologischen Randbedingungen wurden die Bedeutung und die Möglichkeiten akustischer Repräsentationen in kooperativen und hybriden Arbeitsumgebungen untersucht. Zur Realisierung der neuen Konzepte wurden Forderungen und Regeln für die Gestaltung akustischer Repräsentationen aufgestellt und diese dann im Sinne des ‚proof of concept‘ über die Entwicklung von Prototypen für die *i-LAND* Umgebung überprüft. Darüber hinaus wurden die Möglichkeiten der Audiosignalerzeugung und -manipulation auf Standardcomputern untersucht und demonstriert. Mit speziellen Entwicklungen und Implementierungen für die Prototypen konnte gezeigt werden, wie die Verarbeitungsleistung derzeitiger Computer verwendet werden kann, um die Interaktion mit akustischen Signalen zu unterstützen. Um eine Basis zu schaffen, auf der adäquate Lösungen erarbeitet werden konnten, war ein interdisziplinärer Ansatz notwendig, der Erkenntnisse aus verschiedenen Bereichen wie z. B. der Kommunikationstheorie, der Wahrnehmungspsychologie und Psychoakustik sowie der Nachrichtentechnik, der digitalen Signalverarbeitung und der Softwaretechnik berücksichtigte.

Die realisierten Prototypen für neue Formen der Interaktion in der hybriden *i-LAND* Umgebung erweitern in besonderem Maße die realen und virtuellen Qualitäten von hybriden Artefakten mit den Mitteln akustischer Repräsentationen. Die Forderungen und Gestaltungsregeln fassen die wichtigsten Randbedingungen für die Entwicklung von akustischen Repräsentationen in hybriden Umgebungen zusammen. Die Erfüllung der Forderung nach *hörbaren Objekteigenschaften* im anwendungsunabhängigen Teil der Benutzungsschnittstelle wurde z. B. für die Stiftinteraktion an einer großen elektronischen Tafel demonstriert. Für die verschiedenen Roomware-Komponenten wurden unterschiedliche akustische Repräsentationen entwickelt, die in der Lage sind, die jeweils besondere neue Interaktionsform (z. B. an der DynaWall oder das Feedback beim *Passage-Mechanismus*) zu unterstützen. Weitere akustische Repräsentationen haben neue Möglichkeiten für die Vermittlung computerinterner Zustände erschlossen. Zusätzlich wurde die hybride Präsentation von akustischen Signalen vorgestellt, die auf einer Kombination von Kopfhörern und Lautsprechern basiert und damit Anwendungen der *Augmented Reality* im akustischen Medium ermöglicht. Hierzu wurde demonstriert, wie diese Präsentationsform für die Kooperation verwendet werden kann. Die entstandenen Prototypen zeigen, wie das akustische Medium für unterschiedliche Situationen genutzt werden kann, um kohärente und plausible Interaktionen zu erzeugen

und damit die Kooperation von Menschen in zukünftigen hybriden Arbeitsumgebungen zu unterstützen.

Für die Entwicklung dieser Prototypen wurde Software entwickelt, mit der physikalische und abstrakte Modelle akustischer Zusammenhänge für die Anreicherung der Interaktion in Echtzeit möglich wurden. So wurde z. B. für die Interaktion mit einem Stift auf einer Oberfläche ein Geräuschgenerator entworfen, mit dem verdeutlicht werden konnte, wie eine effektive Gestaltung plausibler akustischer Repräsentationen erfolgen kann. Weiterhin wurde die Leistungsfähigkeit von Standardcomputersystemen untersucht. Mit einer beispielhaften Implementierung eines Verfahrens der schnellen Faltung mit langen Impulsantworten wurde die Verwendung von Standardcomputern in interaktiven Anwendungen demonstriert. Es konnte gezeigt werden, dass bei der Verwendung heutiger Standardcomputer eine aufwändige Audiosignalverarbeitung in Echtzeit mit geringer Ausgabeverzögerung möglich ist und für interaktive akustische Repräsentationen benutzt werden kann.

Die Perspektiven, die sich aus den Untersuchungen und Entwicklungen dieser Arbeit ergeben, liegen einerseits direkt in neuen Möglichkeiten zukünftiger Computertechnik und andererseits in der Realisierung neuer Formen der Interaktion zwischen Mensch und Computer. Die aktuellen Trends in der sog. *Post-PC Ära* sind u. a. durch die Mobilität der Benutzer in Verbindung mit dem allgegenwärtigen Zugang zu Informationen und durch den intuitiven Umgang mit dem Computer, z.B. über die Stifteingabe gekennzeichnet. Hierfür stehen einerseits neue Artefakte aus der Konvergenz von Informationstechnologie und Telekommunikation (z. B. PDA-basierte Mobiltelefone), die Eigenschaften besitzen, die in dieser Form bisher noch nicht verfügbar sind. Andererseits entstehen auch neue Geräteklassen, die primär über die Stiftinteraktion bedient werden wie z. B. der *Tablet-PC*. Für diese Geräte sind akustische Repräsentationen für die Interaktion sinnvoll, um dem Benutzer z. B. neue Funktionen der Geräte zu vermitteln. Die vorgestellten Realisierungen akustischer Repräsentationen, die Gestaltungsregeln und die Forderung nach hörbaren Objekteigenschaften, wie sie in dieser Arbeit für hybride Umgebungen entwickelt wurden, können in den zukünftigen kooperativen ‚Informationsräumen‘ helfen, den natürlichen Umgang mit hybriden Artefakten zu unterstützen.

Literaturverzeichnis

- Abowd, G.D. und Mynatt, E.D. (2000). Charting Past, Present, and Future Research in Ubiquitous Computing. *ACM Transactions on Computer-Human Interaction*, 7(1):29–58.
- Beaudouin-Lafon, M. und Gaver, W.W. (1994). ENO: Synthesizing Structured Sound Spaces. In *Seventh Annual Symposium on User Interface Software and Technology (UIST'94)*, 49–57.
- Beaudouin-Lafon, M. und Conversy, S. (1996). Audio illusions for audio feedback. In *Companion Proceedings of ACM CHI 96 Conference on Human Factors in Computing Systems*, Vancouver, Canada, 299–300.
- Bederson, B. (1995). Audio Augmented Reality: A Prototype Automated Tour Guide. In *Companion Proceedings of ACM CHI 95 Conference on Human Factors in Computing Systems*, ACM Press, New York, 210–211.
- Begault, D.R. (1994). *3D Sound for Virtual Reality and Multimedia*. Academic Press, London.
- Begault, D.R. (1999). Auditory and Non-Auditory Factors that Potentially Influence the Virtual Acoustic Imagery. In *Proceedings of the AES 16th International Conference: Spatial Sound Reproduction*, Rovaniemi, Finland, 13-26.
- Bencina, R. und Burk, P. (2001). PortAudio – an Open Source Cross Platform Audio API. In *Proceedings of the International Computer Music Conference*, Havana, Cuba, 263-266.
- Blattner, M.M., Sumikawa, D.A. und Greenberg, R.M. (1989). Earcons and Icons: Their Structure and Common Design Principles. In *Human Computer Interaction*, 4(1):11-44.
- Blauert, J. (1997). *Spatial Hearing: The Psychophysics of Human Sound Localization*. The MIT Press, Cambridge, MA, USA.
- Bregman, A.S. (1990). *Auditory Scene Analysis: The Perceptual Organization of Sound*. The MIT Press, Cambridge, MA, USA.
- Brewster, S. (1994). *Providing a structured method for integrating non-speech audio into human-computer interfaces*. 1994, PhD-Thesis, University of York, U.K.
- Buxton, B. (1989). Introduction to special issue on non-speech audio. *Human-Computer Interaction*, 4(1): 1-9.
- Card, S.K., Moran, T.P. und Newell, A. (1983). *The Psychology of Human-Computer Interaction*. Lawrence Erlbaum, Hillsdale, N.J.
- Cohen, J. (1994a). Monitoring background activities. In: G. Kramer (Hg.). *Auditory display, sonification, audification and auditory interfaces*. Santa Fe Institute, Addison-Wesley, 499-531.

- Cohen, J. (1994b). Out to lunch: further adventures monitoring background activity. In *Proceedings of the International Conference on Auditory Display*, Santa Fe Institute, 15-20.
- Cook, P.R. (1997). Physically Informed Sonic Modeling (PhISM): Synthesis of Percussive Sounds. *Computer Music Journal*, 21(3):38-49.
- Cook, P.R. und Scavone, G.P. (1999). The Synthesis ToolKit (STK), version 2.1. In *Proceedings of the International Computer Music Conference*, Beijing, 164-166.
- Cruz-Neira C., Sandin D., DeFanti T., Kenyon R. und Hart J. (1992). The Cave – Audio Visual Experience Automatic Virtual Environment. *Communications of ACM*, 35(6):64–72.
- Dannenberg, R.B. (1996). The Platform Blues or Looking for Mr. Real Time. *ICMA Array*, 16(1):31-32.
- Darvishi, A., Munteanu, E., Guggiana, V., Schauer, H., Montavalli, M. und Rauterberg, M. (1995). Designing environmental sounds based on the results of interaction between objects in the real world. In: N. Knut, P. Helmersen, D.J. Gilmore, S.A. Arnesen (Hg.). *Human-Computer Interaction: Interact 1995*. London: Chapman & Hall, 38-42.
- Deutsch, D. (1995). *Musical Illusions and Paradoxes*, Audio-CD und Booklet, Philomel Records.
- Dutilleux, P. und Müller-Tomfelde, C. (1999). AML: Architecture and Music Laboratory. In *Proceedings of the AES 16th International Conference: Spatial Sound Reproduction*, Rovaniemi, Finland, 191-206.
- van Doel, K. (1998). *Sound Synthesis for Virtual Reality and Computer Games*. Thesis at the The University of British Columbia.
- van Doel, K., Kry P.G. und Pai, D.K. (2001). Foley Automatic: Physically-based Sound Effects for Interactive Simulation and Animation. In *Proceedings of the ACM SIGGRAPH 2001*, Los Angeles, CA, USA, 537-544.
- Engelen, H. (1998). Sound Design for Consumer Electronics. In *Proceedings Conference on Acoustic Ecology, "Stockholm, Hey Listen!"*, Stockholm, Sweden.
- Dzida, W. (1983). Das IFIP-Modell für Benutzerschnittstellen. *Office Management: Sonderheft Mensch-Maschine-Kommunikation* 31:6-8.
- Eckel, G. (1999). Application for the CyberStage Spatial Sound Server. In *Proceedings of the AES 16th International Conference: Spatial Sound Reproduction*, Rovaniemi, Finland, 478-484.
- Eckel, G. (2001). Immersive Audio-Augmented Environments. In *Proceedings of the 8th Biennial Symposium on Arts and Technology* at Connecticut College, New London, CT, USA.
- Egelmeers, G.P. und Sommen P.C.W. (1994). A new method for efficient convolution in frequency domain by non-uniform partitioning. In *Proceedings EUSIPCO*, Edinburgh, Volume 2, 1030-1033.

- Familant, M.E. und Detweiler, M.C. (1993). Iconic Reference: Evolving Perspectives and an Organizing Framework. *International Journal of Man-Machine Studies*, 39(5):705-728.
- Farina, A., Langhoff, A. und Tronchin, L. (1995). Realization of 'virtual' musical instruments: measurements of the Impulse Response of violins using MLS technique. In *Proceedings of the 2nd International Conference on Acoustics and Musical Research*, Ferrara, Italy, 363-368.
- Feiner, S., MacIntyre, B. und Seligmann, D. (1993). Knowledge-based Augmented Reality. *Communications of the ACM*, 36(7):53-62.
- Finkenzeller, K. (2002). *RFID-Handbuch*. 3. Auflage, Hansa, München.
- Fliege, N.J. (1991). *Systemtheorie*. Teubner, Stuttgart.
- Freed, A. (1993). Clear efficient audio signal processing in ANSI C. *C Users Journal*, 11(9).
- Frigo, M. und Johnson, S. G. (1998). FFTW: An Adaptive Software Architecture for the FFT. In *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing*, 3:1381-1384.
- Gardner, W.G. (1995). Efficient convolution without input-output delay. *Journal of Audio Engineering Society*, 43(3):127-136.
- Gaver, W.W. (1986). Auditory Icons: using sound in computer interfaces. *Human Computer Interaction*, 2(1):167-177.
- Gaver, W.W. (1988). *Everyday listening and auditory icons*. Doctoral Dissertation, University of California, San Diego.
- Gaver, W.W. (1989). The SonicFinder: An Interface That Uses Auditory Icons. In *Human Computer Interaction*, 4(1):67-94.
- Gaver, W.W. (1991a). Sound Support For Collaboration. In *Proceedings of the Second European Conference on Computer-Supported Collaborative Work (E-CSCW'91)*, 293-308.
- Gaver, W.W. (1991b). Technology Affordances. In *Proceedings of the ACM CHI '91 Conference on Human Factors in Computing Systems*, New York, The Association for Computing Machinery, 29-84.
- Gaver, W.W., Smith, R., und O'Shea, T. (1991c). Effective sounds in complex systems: the ARKola simulation. In *Proceedings of the ACM conference on human factors in software (CHI'91)*, 85-90.
- Gaver, W.W. (1993a). Synthesizing auditory icons. In *Proceedings of the INTERCHI '93 - Conference on Human Factors in Computing Systems*, Amsterdam, 228-235.
- Gaver, W.W. (1993b). How do We Hear in the World?: Explorations in Ecological Acoustics. *Ecological Psychology*, 5(4):285-313.
- Gaver, W.W. (1993c). What in the world do we hear? An ecological approach to auditory event perception. *Ecological Psychology*, 5(1):1-29.

- Gaver, W.W. (1997). Auditory Interfaces. In: M.G. Helander, T.K. Landauer, P.V. Prabhu (Hg.), *Handbook of Human-Computer-Interaction*, 2. Edition, Part VI "Multimedia, Video and Voice, Elsevier Science, Amsterdam, 1003-1041.
- Gibson, J.J. (1986). *The Ecological Approach to Visual Perception*. Lawrence Erlbaum Associates, London.
- Guski, R. (2000). *Wahrnehmung. Eine Einführung in die Psychologie der menschlichen Informationsaufnahme*. Kohlhammer, Stuttgart.
- Hermann, T. (1986). *Zur Gestaltung der Mensch-Computer-Interaktion: Systemerklärung als kommunikatives Problem*. Niemeyer, Tübingen.
- Herrmann, T. (2001). Kommunikation und Kooperation. In: G. Schwabe, N. Streitz; R. Unland (Hg.). *CSCW-Kompodium. Lehr- und Handbuch zum computerunterstützten kooperativen Arbeiten*, Springer-Verlag, 15-25.
- Hindus, D., Arons, B., Stifelman, L., Gaver, B., Mynatt, E. und Back, M. (1995). Designing Auditory Interactions for PDAs. In *Proceedings of the Symposium on User Interface Software and Technology (UIST'95)*, 143-146.
- Horbach, U. und Boone, M. M. (1999). Future Transmission and Rendering Formats for Multichannel Sound. In *Proceedings of the AES 16th International Conference: Spatial Sound Reproduction*, Rovaniemi, Finland, 409-418.
- Huron (2000). Lake DSP, <<http://www.lakedsp.com/>>, (24.7.2000).
- IBM Feature Tool, (2001). <http://service.boulder.ibm.com/storage/hddtech/ibmftool_userguide.pdf>, (24.8.2001).
- Ishii, H. und Ullmer, B. (1997). Tangible Bits: Towards seamless Interface between People, Bits and Atoms. In *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI '97)*, 234 – 241.
- Ishii, H., Wisneski, C., Brave, S., Dahley, A., Gorbet, M., Ullmer, B., und Yarin, P. (1998). ambientROOM: Integrating Ambient Media with Architectural Space. In *Video Program of CHI '98*, ACM.
- Jot, J.-M. (1992). *Etude et Realisation d'un Spatialisateur de sons par Modèles Physique et Perceptifs*. PhD-Thesis, Telecom Paris.
- Jot, J.-M. (1999). Real-time spatial processing of sounds for music, multimedia and interactive human-computer interfaces. *ACM Multimedia Systems, Special Issue on Audio and Multimedia*, 7(1):55-69.
- Kammeyer, K.D. und Kroschel, K. (1989). *Digitale Signalverarbeitung*. B.G. Teubner, Stuttgart.
- Karplus, K. und Strong, A. (1983). Digital Synthesis of Plucked-String and Drum Timbres, *Computer Music Journal*, 7(2):43-55.
- Kendall, G.S. (1995a). A 3-D Sound Primer: Directional Hearing and Stereo Reproduction, *Computer Music Journal*, 19(4):23-46.
- Kendall, G.S. (1995b). The Decorrelation of Audio Signals and Its Impact on Spatial Imagery, *Computer Music Journal*, 19 (4):72-87.

- Kleiner, K., Dalenbäck, B.-I. und Svensson P. (1993). Auralisation - an Overview. *Journal of Audio Engineering Society*, 41(11):861-875.
- Konomi, S., Müller-Tomfelde, C., und Streitz, N. (1999). Passage: Physical transportation of digital information in cooperative buildings. In: N. Streitz, J. Siegel, V. Hartkopf, and S. Konomi (Hg.). *Proceedings of the Second International Workshop on Cooperative Buildings (CoBuild'99)*, LNCS 1670, Springer, Heidelberg, Germany, 45–54.
- Kramer, G. (1994). An Introduction to Auditory Display. In: G. Kramer (Hg.). *Auditory display, sonification, audification and auditory interfaces*. Santa Fe Institute, Addison-Wesley, 1-77.
- Kraut, R.E., Fish, R.S., Root, R.W. und Chalfonte, B.L. (1990). Informal Communication in Organizations: Form, Function, and Technology. In: R. Baecker: *Readings in Groupware and Computer-Supported Cooperative Work*, Morgan Kaufman, 145-199.
- Kubovy, M. und van Valkenburg, D. (2001). Auditory and visual objects, *Cognition*, 80:97-126.
- Lindemann, E., Dechelle, F., Smith, B. und Starkier, M. (1991). The Architecture of the IRCAM Musical Workstation. *Computer Music Journal*, 15(3):41-49.
- Marple, S.L. (1987). *Digital spectral analysis with applications*. Prentice-Hall Englewood Cliffs, NJ, USA
- Max/MSP (1997). Cycling 74, <<http://www.cycling74.com>>, (9.10.2001).
- MacMillan, K., Droettboom M. und Fujinaga, I. (2001). Audio Latency Measurements of Desktop Operating Systems. In *Proceedings of International Computer Music Conference*, Havana, Cuba, 259-262.
- Milgram, P. und Kishino, F. (1994). A Taxonomy of Mixed Reality Visual Displays, *IEICE Transactions on Information Systems*, Vol E77-D (12), 1321-1329.
- Moorer, J.A. (1987). About This Reverberation Business, Chapter Perception and Digital Signal Processing, *Foundations of Computer Music*. The MIT Press, Cambridge, MA, USA., 605-639.
- Moorer, J.A. (2000). Audio in the New Millennium, *Journal of Audio Engineering Society*, 48(5):490-498.
- Moran, T.P. (1981). The command language grammar: a representation for the user interface of interactive computer systems. In *International Journal of Man-Machine Studies*, 15,3-51.
- Müller-Tomfelde, C. und Reischl, W. (1998). Communication Chairs: Examples of Mobile Roomware Components. In *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI '98)*, Los Angeles, USA, 267-268.
- Müller-Tomfelde, C. (1999). Low Latency convolution for real time application. In *Proceedings of the AES 16th International Conference: Spatial Sound Reproduction*, Rovaniemi, Finland, 454-460.

- Müller-Tomfelde, C. und Steiner, S. (2001). Audio-Enhanced Collaboration at an Interactive Electronic Whiteboard. In *Proceedings of 7th International Conference on Auditory Display (ICAD01)*, Espoo, Finland, 267-271.
- Müller-Tomfelde, C. (2001). Time-Varying Filter in Non-Uniform Block Convolution. In *Proceedings of the COST G-6 Conference on Digital Audio Effects (DAFX-01)*, Limerick, Ireland, 105-109.
- Müller-Tomfelde, C. und Münch, T. (2001). Modeling and Sonifying Pen Strokes on Surfaces. In *Proceedings of the COST G-6 Conference on Digital Audio Effects (DAFX-01)*, Limerick, Ireland, 175-179.
- Müller-Tomfelde, C. (2002a). Hybrid Sound Reproduction in Augmented Reality. In *Proceedings of the AES 22th International Conference: Virtual, Synthetic, and Entertainment Audio*, Espoo, Finland, 58-63.
- Müller-Tomfelde, C. (2002b). Sound Effects for a Silent Computer System. In *Proceedings of the COST G-6 Conference on Digital Audio Effects (DAFX-02)*, Hamburg, Germany, 227-232.
- Mynatt, E.D., Back, M., Want, R., Baer, M. und Ellis, J. B. (1998). Designing audio aura. In *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI '98)*, Los Angeles, USA, 566-573.
- Nigay, L. und Coutaz, J. (1993). A design space for multimodal systems - concurrent processing and data fusion. In *Proceedings of the INTERCHI '93 - Conference on Human Factors in Computing Systems*, Amsterdam, 172-178.
- Norman, D.A. (1986). Cognitive Engineering. In: D. A. Norman and S. W. Draper (Hg.). *User Centered System Design*, Erlbaum, Hillsdale NJ, 31-61.
- Norman, D. A. (1988). *The psychology of everyday things*. Basic Books, New York.
- Pöppel, E. (1997). A hierarchical model of temporal perception. *Trends in Cognitive Science* 1(2):56-61.
- PortAudio (2001). <<http://www.portaudio.com/>>, (28.5.2001).
- Project Aura (2001). Carnegie Mellon University, <<http://www-2.cs.cmu.edu/~aura/>>, (4.5.2001).
- Project Oxygen (2001). MIT Laboratory for Computer Science, <<http://oxygen.lcs.mit.edu/>>, (29.3.2001).
- Puckette, M. (1991a). FTS: A real-time monitor for multiprocessor music synthesis. *Computer Music Journal*, 15(3), 58-67.
- Puckette, M. (1991b). Combining event and signal processing in the Max graphical programming environment. *Computer Music Journal*, 15(3), 68-77.
- Puckette, M. (1996). Pure Data: another integrated computer music environment. In *Proceedings of the Second Intercollege Computer Music Concerts*, Tachikawa, 31-41.
- Pure Data (2001). <<http://www.pure-data.org/>>, (21.11.2001).

- Rasmussen, J. (1986). *Information Processing and Human-Machine Interaction. An Approach to Cognitive Engineering*. Elsevier Science, New York.
- Rauterberg, M., Motavalli, M., Darvishi, A. und Schauer, H. (1994). Automatic sound generation for spherical objects hitting straight beams based on physical models. In *Proceedings of ED-MEDIA 94-Conference on Educational Multimedia and Hypermedia Vancouver*, BC, Canada, 469-473.
- Rauterberg, M. (1995). Human Information Processing in Man-Machine Interaction. In: A. Grieco, G. Molteni, B. Piccoli and E. Occhipinti (Hg.). *Work with Display Units 94*, Elsevier Science B.V., 221-226.
- Roads, C. (1996). *The Computer Music Tutorial*, 3. Edition, The MIT Press, Cambridge, MA, USA.
- Robertson, G.G. (1998). <<http://research.microsoft.com/~ggr/pui98.htm>>, (10.4.2002).
- Rocchesso, D. (1998). Fractionally-Addressed Delay Lines. In *Proceedings of the COST G-6 Conference on Digital Audio Effects (DAFX-98)*, Barcelona, Spain, 40-43.
- Rocchesso, D. (2001). Acoustic cues for 3-d shape information. In *Proceedings of 7th International Conference on Auditory Display (ICAD01)*, Espoo, Finland, 180-183.
- Sawhney, N. und Schmandt, C. (1999). Nomadic radio: Scaleable and contextual notification for wearable audio messaging. In *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI '99)*, New York, NY, USA, 96-103.
- Schomaker, L., Nijtmans, J., Camurri, A., Lavagetto, F., Morasso, P., Benoit, C., Guiard-Marigny, T., Le Goff, B., Robert-Ribes, J., Adjoudani, A., Defee, I., Münch, S., Hartung, K. und Blauert, J. (1995). *A Taxonomy of Multimodal Interaction in the Human Information Processing System*. A Report of the Esprit BRA Project 8579 MIAMI, WP1.
- Schröder, M.R. (1970). Digital Simulation of Sound Transmission in Reverberant Spaces. In *Journal of the Acoustic Society of America*, 47:424-431.
- Schwabe, G. (2001). Theorie zur Mediennutzung bei der Gruppenarbeit. In: G. Schwabe; N. Streitz; R. Unland (Hg.). *CSCW-Kompodium. Lehr- und Handbuch zum computerunterstützten kooperativen Arbeiten*, Springer-Verlag.
- Schwarz, H.P. (1997). *Medien-Kunst-Geschichte*. Medienmuseum ZKM | Zentrum für Kunst und Medientechnologie, Karlsruhe. Prestel, München.
- Serra, X. und Smith, J.O. (1990). Spectral modeling synthesis: A sound analysis/synthesis system based on a deterministic plus stochastic decomposition. *Computer Music Journal*, 14(4):12-24.
- Shannon, C.E. und Weaver, W. (1949). *The Mathematical Theory of Communication*. Urbana (Illinois): The University of Illinois Press.
- Shneiderman, B. (1992). *Designing the User Interface - Strategies for Effective Human-Computer Interaction*, 2nd Edition, Addison-Wesley.

- Siira, J. und Pai, D.K. (1996). Haptic Textures – A Stochastic Approach. In *Proceedings of the International Conference on Robotics and Automation*, Minneapolis MN, 557-562.
- Smith, J.O. (1991). Viewpoints on the History of Digital Synthesis. In *Proceedings of International Computer Music Conference (ICMC-91)*, Montreal, Canada, 1-10.
- Stautner, J. und Puckette, M. (1982). Designing multi-channel reverberators. *Computer Music Journal*, 6(1):52-65.
- Steinmetz, R. (2000). *Multimedia-Technologie, Grundlagen Komponenten und Systeme* 3. Auflage, Springer.
- Stockham, T.G. (1966), High-speed convolution and correlation. In *Proceedings of the AFIPS Spring Joint*, Spartan Books, 28:229 - 233.
- Streitz, N.A. (1985). Die Rolle von mentalen und konzeptuellen Modellen in der Mensch-Computer-Interaktion: Konsequenzen für die Software-Ergonomie? In: H.-J. Bullinger (Hg.), *Software-Ergonomie '85. Mensch-Computer- Interaktion*, B.G. Teubner, Stuttgart, 281-292.
- Streitz, N.A., Geißler, J., Holmer, T., Konomi, S., Müller-Tomfelde, C., Reischl, W., Rexroth, P., Seitz, P. und Steinmetz, R. (1999). An interactive Landscape for Creativity and Innovation. In *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI '99)*, Pittsburgh, Pennsylvania, USA, ACM Press, New York, 120-127.
- Streitz, N.A., Tandler, P., Müller-Tomfelde, C., Konomi, S. (2001) Roomware: Towards the Next Generation of Human-Computer Interaction based on an Integrated Design of Real and Virtual Worlds. In: J. A. Carroll (Hg.): *Human-Computer Interaction in the New Millennium*, Addison Wesley, 553-578.
- Takala, T. und Hahn, J. (1992). Sound rendering. In *Proc. SIGGRAPH 92, ACM Computer Graphics*, 26(2): 211–220.
- Tandler, P. (2000). Architecture of Beach: The Software Infrastructure for Roomware Environments. In: *CSCW 2000: Workshop on Shared Environments to Support Face-to-Face Collaboration*, Philadelphia, Pennsylvania, USA.
- Tandler, P. (2001). Software Infrastructure for a Ubiquitous-Computing Environment Supporting Collaboration with Multiple Single- and Multi-User Devices. In *Proceedings of UbiComp'01* (Atlanta, USA), Lecture Notes in Computer Science, Springer, Heidelberg, Germany, 96-115.
- Tandler P., Prante, T., Müller-Tomfelde, C., Streitz, N.A. und Steinmetz, R. (2001). ConnecTables: Dynamic Coupling of Displays for the Flexible Creation of Shared Workspaces. In *Proceedings of the 14th Annual ACM Symposium on User Interface Software and Technology (UIST'01)*, Orlando, Florida, USA, CHI Letters 3(2), ACM Press, 11-20.
- Terhardt, E. (1998). *Akustische Kommunikation*, Springer.
- Turk, M. und Robertson, G. (2000). Perceptual User Interfaces. *Communications of the ACM*, 43(3):33-34.

- Ullmer, B. und Ishii, H. (2000). Emerging frameworks for tangible user interfaces. *IBM Systems Journal*, 39(3 u. 4): 915-931.
- Virilio, P. (1990). *Rasender Stillstand*, Carl Hansa Verlag.
- Want, R., Schilit, B., Norman A., Gold R., Goldberg D., Petersen K., Ellis J. und Weiser, M. (1995). An Overview of the ParcTab Ubiquitous Computing Experiment, *IEEE Personal Communications*, 2(6):28-43.
- Want, R., Hopper, A., Falcao, V. und Gibbons, J. (1992). The Active Badge Location System. *ACM Transactions on Information Systems*, 10(1):91-102.
- Warren, W. H. und Verbrugge, R. R. (1984). Auditory perception of breaking and bouncing events: a case study in ecological acoustics. *Journal of Experimental Psychology: Human Perception and Performance*, 10(5):704-712.
- Weiser, M. (1991). The computer for the twenty-first century. *Scientific American*, 265(3), 94-104.
- Weiser, M. (1996). Designing Calm Technology. *PowerGrid Journal*, v101, July 1996, <<http://powergrid.electriciti.com/1.01>>, (23.2.2000).
- Weizsäcker von, E. (1974). Erstmaligkeit und Bestätigung als Komponenten der pragmatischen Information. In: E. von Weizsäcker (Hg.): *Offene Systeme, Band I. Beiträge zur Zeitstruktur von Information, Entropie und Evolution*. Klett, 82-113.
- Wellner, P. (1993). Interacting with paper on the DigitalDesk. *Communications of the ACM*. 36(7):87-96.
- Wenzel, E.M., Miller, J.D. und Abel, J.S. (2000). A software-based system for interactive spatial sound synthesis. In *Proceedings of 7th International Conference on Auditory Display (ICAD00)*, Atlanta, GA, USA.
- Wyse, L. und Kellock, P. (1999). Embedding interactive sounds in multimedia applications. *ACM Multimedia Systems, Special Issue on Audio and Multimedia*. 7(1):55-69.
- Wrightson, K. (2000). An Introduction to Acoustic Ecology. *Soundscape*, 1(1):12-13.
- Zölzer, U. (1997). *Digital Audio Signal Processing*. J. Wiley & Sons, Chichester.

Anhang

A.1 Inhalt der Daten-CD

Die beigefügte Daten-CD enthält vier beispielhafte Demonstrationen der beschriebenen Prototypen der akustischen Darstellungen aus Kapitel 6. Die kurzen Videosequenzen verfügen über eine Audiospur, die direkt vom Ausgang der Klangerzeugung der jeweiligen akustischen Darstellung aufgezeichnet wurde. Alle weiteren Geräusche der Umgebung und Stimmen der Personen wurden nicht aufgezeichnet. Auch wenn die Wahrnehmung bei der Interaktion nicht der entspricht, die beim Hören und Betrachten der Audio/ Videosequenzen entsteht, sollte doch damit versucht werden einen sinnlichen Eindruck der jeweiligen Interaktion zu vermitteln. Der Betrachter wird in die Rolle des Beobachters der Interaktion von Personen versetzt, ohne selbst zu interagieren.

Audio-Unterstützung der Interaktion und Kooperation an der DynaWall

In kooperativen Arbeitssituationen an der DynaWall werden die verschiedenen Formen der akustischen Repräsentationen für die Tafeloberfläche in der hybriden i-LAND Umgebung demonstriert. Mehrere Personen arbeiten in einer Kleingruppe an der DynaWall. Neben den grundlegenden Formen der Anreicherung der Interaktion mit einer akustischen Darstellung wird auch die Gestenmelodie demonstriert.

Datei: DynaWall-Interaktion.avi, 170 s

Erweiterte akustische Modellierung der Stiftinteraktion

An einem Graphiktablett wird die Wirkung der erweiterten Modellierung der Stiftinteraktion demonstriert (siehe auch Abb. 26). Die verschiedenen Farbbereiche auf der Oberfläche repräsentieren verschiedene Kombinationen von Anteilen der akustischen Oberflächentexturen. Der dabei verwendete spezielle Stift hat eine Schreibspitze und eine Radiergummiseite, die bei der akustischen Darstellung unterschiedlich eingesetzt werden. Während bei der Interaktion mit der normalen Stiftspitze typische Geräusche eines Boardmarkers ertönen, wird mit der Radiergummiseite des Stiftes lediglich die Meso-Textur des Klanggenerators stimuliert.

Datei: Stiftinteraktion.avi, 73 s

Audiofeedback für den Passage-Mechanismus

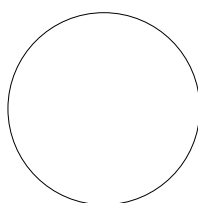
In der Demonstrationssequenz wird mit einem roten Klotz ein elektronisches Dokument von der DynaWall zum InteracTable transportiert. Dabei wird die elektronische Waage zur Identifikation des Objektes verwendet. Für die akustische Darstellung ist das abstrakte Schema in der Demonstration gewählt, das dem Benutzer den Prozess des Mechanismus hörbar macht (siehe Abschnitt 6.5.3).

Datei: Passage-Feedback.avi, 61 s

Akustischer Computersystemmonitor

An der Roomware-Komponente ConnecTable wird die Interaktion des Benutzers durch die Anreicherung mit Geräuschen unterstützt. Der ConnecTable ist ein hybrides Artefakt, das so konzipiert wurde, dass im Betrieb keine Störgeräusche z. B. aufgrund von drehenden Teilen produziert werden. Wie in Abschnitt 6.6 beschrieben, wird daher die CPU-Auslastung und die Größe des Datentransports über das Netzwerk in hörbare Geräusche über die Audioausgabe umgesetzt. In der Demonstration verwendet der Mitarbeiter eine Standardanwendung, bei der sowohl Netzwerkaktivität als auch CPU Leistung je nach Interaktion gefordert und damit auch hörbar wird. Diese Darstellung interferiert mit den im Betriebssystem einstellbaren Geräuschen z. B. beim Anklicken von Schaltflächen.

Datei: Akustischer-Monitor.avi, 59 s.



A.2 Abkürzungen

AES	Audio Engineering Society
ANSI	American National Standards Institute
BEACH	Basic Environment for Active Collaboration with Hypermedia
CCRMA	Center for Computer Research in Music and Acoustics
CPU	Central Processing Unit
DSP	Digital Signal Processor
EAR	Environmental Audio Reminder
EBU	European Broadcast Union
FIR	Finite Impulse Response
FM	Frequenzmodulation
FFT	Fast Fourier Transformation
GUI	Graphical User Interface
HMD	Head Mounted Display
HRTF	Head Related Transfer Function
ICAD	International Conference on Auditory Display
IFFT	Inverse Fast Fourier Transformation
IFIP	International Federation for Information Processing
IID	Interaural Intensity Difference
IIR	Infinite Impulse Response
IMW	IRCAM Musical Workstation
IRCAM	Institut de Recherche et Coordination Acoustique/ Musique
ITD	Interaural Time Delay
KARMA	Knowledge-based Augmented Reality Maintenance Assistance
LAN	Local Area Network
LFO	Low Frequency Oscillator
LPC	Linear Prediction Coding
LTI	Linear Time-Invariant
LTV	Linear Time-Variant
MIDI	Musical Instrument Digital Interface
MSP	Max Signal Processing

OOP	Objektorientierte Programmierung
PARC	Palo Alto Research Center
PCA	Principal Component Analysis
PDA	Personal Digital Assistant
PDH	Performance Data Helper
RFID	Radio Frequency Identification
PHM	Physical Modeling
RISC	Reduced Instruction Set Computer
SIMD	Single Instruction Multiple Data
SPDIF	Sony Philips Digital Interface
STK	Synthesis ToolKit
TUI	Tangible User Interface

A.3 Publikationsliste im Kontext der Dissertation

- Müller-Tomfelde, C. und Dutilleux, P. (1996). Eine Installation zur Raumakustik für das Medienmuseum des ZKM. *19. Tonmeistertagung*, Bildungswerk des Verbandes Deutscher Tonmeister (VDT), Berlin, 481-488.
- Müller-Tomfelde, C. und Reischl, W. (1998). Communication Chairs: Examples of Mobile Roomware Components. In *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI '98)*, Los Angeles, USA, 267-268.
- Müller-Tomfelde, C. (1999). Low Latency convolution for real time application. In *Proceedings of the AES 16th International Conference: Spatial Sound Reproduction*, Rovaniemi, Finland, 454-460.
- Dutilleux, P. und Müller-Tomfelde, C. (1999). AML: Architecture and Music Laboratory. In *Proceedings of the AES 16th International Conference: Spatial Sound Reproduction*, Rovaniemi, Finland, 191-206.
- Streitz, N.A., Geißler, J., Holmer, T., Konomi, S., Müller-Tomfelde, C., Reischl, W., Rexroth, P., Seitz, P. und Steinmetz, R. (1999). An interactive Landscape for Creativity and Innovation. In *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI '99)*, Pittsburgh, Pennsylvania, USA, ACM Press, New York, 120-127.
- Konomi, S., Müller-Tomfelde, C., und Streitz, N. (1999). Passage: Physical transportation of digital information in cooperative buildings. In: N. Streitz, J. Siegel, V. Hartkopf, and S. Konomi (Hg.). *Proceedings of the Second International Workshop on Cooperative Buildings (CoBuild'99)*, LNCS 1670, Springer, Heidelberg, Germany, 45-54.
- Müller-Tomfelde, C. und Steiner, S. (2001). Audio-Enhanced Collaboration at an Interactive Electronic Whiteboard. In *Proceedings of 7th International Conference on Auditory Display (ICAD01)*, Espoo, Finland, 267-271.
- Streitz, N.A., Tandler, P., Müller-Tomfelde, C., Konomi, S. (2001) Roomware: Towards the Next Generation of Human-Computer Interaction based on an Integrated Design of Real and Virtual Worlds. In: J. A. Carroll (Hg.): *Human-Computer Interaction in the New Millennium*, Addison Wesley, 553-578.
- Tandler P., Prante, T., Müller-Tomfelde, C., Streitz, N.A. und Steinmetz, R. (2001). ConnecTables: Dynamic Coupling of Displays for the Flexible Creation of Shared Workspaces. In *Proceedings of the 14th Annual ACM Symposium on User Interface Software and Technology (UIST'01)*, Orlando, Florida, USA, CHI Letters 3(2), ACM Press, 11-20.
- Müller-Tomfelde, C. (2001). Time-Varying Filter in Non-Uniform Block Convolution. In *Proceedings of the COST G-6 Conference on Digital Audio Effects (DAFX-01)*, Limerick, Ireland, 105-109.
- Müller-Tomfelde, C. und Münch, T. (2001). Modeling and Sonifying Pen Strokes on Surfaces. In *Proceedings of the COST G-6 Conference on Digital Audio Effects (DAFX-01)*, Limerick, Ireland, 175-179.

- Müller-Tomfelde, C. (2002a). Hybrid Sound Reproduction in Augmented Reality. In *Proceedings of the AES 22th International Conference: Virtual, Synthetic, and Entertainment Audio*, Espoo, Finland, 58-63.
- Müller-Tomfelde, C. (2002b). Sound Effects for a Silent Computer System. In *Proceedings of the COST G-6 Conference on Digital Audio Effects (DAFX-02)*, Hamburg, Germany, 227-232.

A.4 Lebenslauf

Christian Müller-Tomfelde

Geboren am 24. Februar 1965 in Hamburg

Schulbildung:

1971-1975	Maler-Becker Schule, Mainz
1975-1984	Gymnasium Mainz-Gonsenheim

Wehrdienst:

1984-1985	Bundesmarine Wilhelmshaven
-----------	----------------------------

Hochschulausbildung:

1986-1989	Studium der Elektrotechnik an der Technischen Universität Darmstadt
1989-1993	Fortsetzung und Abschluss des Studiums an der Technischen Universität Hamburg-Harburg

Beruf:

1994-1997	Forschungs- und Entwicklungstätigkeit als freier Mitarbeiter am Zentrum für Kunst und Medientechnologie ZKM, Karlsruhe
1997-2002	Wissenschaftlicher Mitarbeiter am Fraunhofer Institut für Integrierte Publikations- und Informationssysteme, Darmstadt

